An Elementary Treatise

on

FREQUENCY CURVES

and their Application in the Analysis

of

Death Curves and Life Tables

by

Arne Fisher.

Translated from the Danish

by

E. A. Vigfusson.

With an Introduction

by

Raymond Pearl,
Professor of Biometry and Vital Statistics
Johns Hopkins University, Baltimore.

American Edition.
New York.
THE MACMILLAN COMPANY.
1922.

Printed by Bianco Luno, Copenhagen.

INTRODUCTION

The fact that actuarial science is fundamentally a branch of biology rather than of mathematics is overlooked far more generally than ought to be the case. Most people, even those of education and wide culture, are inclined to look upon an actuary as a particularly crabbed, narrow, and intellectually dusty kind of mathematician. In reality his subject is one of the liveliest in the whole domain of biology, and none surpasses it in its practical interest and importance to mankind. Because, what the actuary is, or at least should be, trying always to formulate more and more definitely are the laws which determine the duration of human life. Why the actuary in fact is too often intellectually but little more than a sort of glorified computer, is really only the result of a defect in the teaching of biology in our colleges and universities. It has only lately come to be recognized anywhere that a biologist needed a substantial foundation in mathematics in order successfully to practice a biological profession. It is not too rash a prediction to say that presently the time is coming when no important actuarial post will be held by a mathematician who knows little or no biology. The vigor and originality of his biological outlook will be valued as highly as the rigidity of his mathematical substructure now is.

The thing which chiefly makes this book by my friend Arne Fisher notable, lies, in a broad sense, in the fact that it is a highly original and absolutely novel essay in general biology. The language is to a considerable extent mathematical, to be sure, but the subject matter, the mode of logical approach, and the significant conclusion — all these are pure biology. Unfortunately many biologists will not be able to appreciate its significance, or even to read it intelligently. But this is their loss, and at the same time an exposure of the dire poverty of their intellectual equipment for dealing with the problems of their science.

There are two broad features of Fisher's work which want emphasis. The first is the successful construction of a life table from a knowledge of deaths alone. That the construction is successful his results set forth in this book abundantly demonstrate. To have done this is a mathematical and actuarial achievement of the first rank. It may fairly be regarded as *fundamentally* the most significant advance in actuarial theory since Halley. It opens out wonderful possibilities of research on the laws of mortality, in directions which have hitherto been wholly impossible of attack. The criterion by which the significance of a new technique in any branch of science is evaluated, is just this of the degree to which it opens up new fields of research. By this criterion Fisher's work stands in a high and secure position.

But of vastly more significance considered purely as an intellectual achievement is his discovery of the fundamental biological law relating the several causes of death to each other, which made the technical accomplishment possible. More than one accepted

text book on vital statistics has scornfully instructed its readers that no good whatever could come from any tabulation or study of death ratios; that they must be avoided as the pestilence by any statistician who would be orthodox. But orthodoxy and discovery are as incompatible intellectually as oil and water are physically, a cosmic law often overlooked by our "safe and sane" scientific gentry. This book is an outstanding demonstration that this law is still in operation. Fisher has had the temerity to study the ratios of deaths from one cause or group of causes to those from another group, or to all causes together, and has discovered that there abides a real and hitherto unsuspected lawfulness in these ratios. Here again his pioneer work opens out alluring vistas to the thoughtful biometrican.

Altogether we of America are to be warmly congratulated that this brilliant Danish mathematical biologist has chosen to come and live with us.

Baltimore, November 1921.

Raymond Pearl.

AUTHOR'S PREFACE

The classical method of measuring mortality rests essentially upon the fundamental principles first enunciated by the British astronomer, Halley, in his construction of the famous Breslau Life Table. Since the time of Halley this method has been so thoroughly investigated and has been perfected to such an extent that new developments along this line cannot be expected. Any improvements on the original principles of Halley are after all nothing but refinements in graduating methods; and even in this line it appears that the limit of further perfection has been reached.

Halley's method, which is purely empirical in scope and principle, rests primarily upon the knowledge of the number of persons exposed to risk at various ages and the correlated number of deaths among such exposures. In all cases where such information is at hand the old and tried method meets all requirements to our full satisfaction; and it would appear superfluous to try to supplant it with fundamentally different principles.

In presenting the new method outlined in this little book I wish to state most emphatically that it has never been my intention to try to supersede the conventional methods of constructon of mortality tables wherever such methods are applicable. My proposed method is only a supplement to the former

tools of statisticians and actuaries, and aims to utilize numerous statistical materials to which the older system of Halley is not applicable. The idea, whether it is new or not, meets in reality a very frequent need in mortality investigations. It is a well known fact that in the determination of certain statistical ratios, it is easier to determine the numerator than the denominator, as for instance in life or sickness assurance, where the losses can be ascertained with a very close degree of accuracy, while the collection of persons exposed to risk at various ages is often difficult to obtain. Similar remarks hold true in the case of numerous statistical summaries of mortuary records as published in most government reports on vital statistics. The desire to utilize this enormous statistical material was what led me to try the proposed method.

In principle the plan is fundamentally different from that of the empirical method of Halley, inasmuch as I have attempted to substitute the inductive principle for that of pure empiricism.

In the first place, I consider the d_x curve, or the number of deaths by attained ages among the survivors of an original cohort of say 1,000,000 entrants at age 10, as being generated as a compound curve of a limited number (say 8 or less) of subsidiary component curves of either the Laplacean-Charlier or Poisson-Charlier type.

The method of induction now consists in determining the constants or parameters of these subsidiary curves. These parameters fall into two separate categories:—

A. The statistical characteristics or semi-invariants which determine the relative frequency distribu-

tion by attained age at death, as expressed by the mean, the dispersion, the skewness and the excess of each subsidiary or component curve.

B. The areas of each subsidiary or component curve.

The working hypothesis which I have put forward is that *the relative frequency distribution of deaths by attained ages, classified according to a limited number of groups (generally 8 or less) of causes of death among the survivors of the original cohort of entrants, tend to cluster around certain ages in such a way that it is possible from biological considerations to estimate in practice with a sufficiently close degree of approximation the statistical characteristics or semi-invariants of the relative frequency distributions of the component curves, corresponding to a previously chosen classification of causes of death (into 8 or less subsidiary groups).*

This implies briefly that I suppose it is possible from biological considerations to select *a priori* the statistical characteristics of the category as mentioned above under A.

Once this hypothesis is accepted as a true supposition, the areas of each of the component curves can be determined by purely deductive methods (as for instance the method of least squares) from the observed values of the proportionate death ratios $R_B(x)$ ($x = 10, 11, 12, \ldots\ldots 100$; $B = \text{I, II, III, } \ldots\ldots$) corresponding to the groups of causes of death.

Thus the parameters as determined in this manner exhaust the given statistical material, i.e. the observed proportionate death ratios $R_B(x)$. A mere addition of the subsidiary or component curves

gives us then the compound d_x curve from which it is an easy task to find the functions, l_x and q_x.

The scheme as we have briefly outlined it above is, therefore, not a cut-and-dried doctrine or a sort of "mathematical alchemy" as some of my critics have implied. Nor is it an authoritative or infallible dogma. The keystone upon which its success depends is merely a working hypothesis; i.e. a temporary or preliminary supposition. I suppose something to be true and try to ascertain whether, in the light of that supposed truth, certain facts fit together better than they do with any other supposition hitherto tried.

The validity of the working hypothesis must, in my opinion, be proved or disproved either by independent methods and principles of construction of mortality tables, such as for instance the empirical principle of Halley, hitherto exclusively used by the actuaries, *or through additional biological studies.* [1]

[1] The biological basis of Mr Fisher's working hypothesis, which is of far greater importance than the purely ancillary mathematical deduction, has apparently been overlooked by many of his American critics, such as Little, Thompson and Carver. Dr. Carver in the *Proceedings of the Casualty Actuarial Society of America* (Vol. VI, page 357) remarks that "if we can construct a table from death alone as in *Proc.* Vol. IV, and by dividing these deaths by q_x, determine the unenumerated population — why not the converse?"

The answer to this remark is obvious. In the case of mortuary records, Fisher considered two different and distinct attributes, namely 1) the purely *quantitative* attribute of attained age at death, and 2) the purely *biological* attribute of cause of death, which in conjunction with the working hypothesis to a certain extent aims to replace the unknown exposures. If we were to follow Dr. Carver's facetious suggestion and, to use his phrase, "go the proposed plan one better by using enumerated populations only", we should, however, encounter a statistical series with the single attribute of attained age only, but no second attribute corresponding to that of the biological factor of the cause of death. Criticisms

In the meantime I feel justified in presenting to my readers the practical results obtained by this method, which although perhaps not unimpeachable in respect to mathematical rigour, nevertheless in my opinion offers a means to attack a vast bulk of collected statistical data against which our former actuarial tools proved useless. The celebrated Russian mathematician Tchebycheff, once made a remark to the effect that in the antique past the Gods proposed certain problems to be solved by man, later on the problems were presented by halfgods and great men, while now dire necessity forces us to seek some solution to numerous practical problems connected with our daily conduct. The problem towards which I have made an attempt to offer a sort of solution in the present little essay is one of these numerous problems of dire necessity mentioned by Tchebycheff, and I hope that my work along this line, imperfect as it is, may nevertheless prove a beginning towards more improved methods in the same direction.

In conclusion I wish to extend my thanks to a number of friends and colleagues both in America and Europe and Japan who have kept on encouraging me in my work along these lines in spite of much adverse criticism from certain statistical and actuarial circles. I wish in this connection to thank Mr. F. L. Hoffman, Statistician of the Prudential Insurance Company, for permitting me to apply the method to various collections of mortuary records while working as a computer in his department. My thanks are also

---

of the sort of Dr. Carver's brings to light the fundamentally different principles applied by Mr. Fisher in sharp contradistinction to the purely empirical methods of the orthodox actuary and statistician.

Translator.

due to Mr. E. A. Vigfusson for making the translation from my rough Danish notes. If the resulting English is perhaps open to criticsm, I beg to remind the reader that my original manuscript was written in Danish and translated into English by an Icelander, while the composition and proof reading was done by a Copenhagen firm.

To Professor Glover of the University of Michigan I also wish to extend my thanks for inviting me to deliver a series of lectures on the construction of mortality tables before his classes in actuarial methods during the month of March 1919. This invitation afforded me the first opportunity to bring the proposed method before a professional body of statistical readers.

Last but not least I desire to acknowledge my obligations to Professor Pearl whose introductory note I consider the strongest part of the book. In these departments of knowledge the appreciation of one's peers is after all the only real reward one can possibly expect. The fact that this eminent biologist has recognized that the nucleus of the whole problem is of a purely biological nature, and that the mathematical analysis is merely ancillary, is particularly pleasing to me, because it represents my own view in this particular matter.

p. t. Newark, U. S. A., November 1921.

Arne Fisher.

TRANSLATOR'S PREFACE

During the spring of 1919 the attention of the present writer was called to a brief paper entitled *Note on the Construction of Mortality Tables by means of Compound Frequency Curves* by the Danish statistician, Mr. Arne Fisher. The novelty and originality of this paper impressed me to such an extent that I became desirous of obtaining more detailed information about the process than that which necessarily was contained in the above summary note, originally printed in the *Proceedings of the Casualty and Acturial Society of America.*

I wrote therefore to Mr. Fisher and inquired whether he intended to publish any further studies on this subject. From his reply I learned that he had delivered a series of lectures on this very topic before Professor Glover's insurance classes at the University of Michigan during the month of March 1919, but that the proposed method had been met with such captious opposition in certain actuarial circles that he had decided to abandon the plan of publishing anything further on the subject and had even destroyed the English notes prepared for the Michigan lectures.

In the meantime the proposed scheme had received considerable attention in actuarial circles in Europe and Japan and several highly commendatory

reviews had appeared in the English and Continental insurance periodicals and various scientific journals, notably the *Journal of the Royal Statistical Society* and the *Bulletin de l'Association des Actuaires Suisses.* The proposed method seemed indeed so novel and unique that I could not help feeling that it deserved a better fate than that of being forgotten. I suggested therefore to Mr. Fisher that he prepare a new manuscript. But unfortunately his time did not allow this. He consented, however, to turn over to me his original Danish notes on the subject from which he had prepared his Michigan lectures and permitted me to make an English translation for the *Scandinavian Insurance Magazine.* I gladly availed myself of this opportunity to bring this fundamental work before an international body of readers and started on the translation in the summer of 1919.

At the same time Mr. Fisher decided to put the proposed method and working hypothesis to a very severe test, which would meet even the most stringent requirements of some of his critics and their contention that the method would fail in the case of a rapidly changing population group. For this purpose he selected a series of statistical data contained in the annual reports and statements of a number of the leading Japanese Life Assurance Offices, relating to their mortuary records for the four year period from 1914—1917. More than 35,000 records of male lives, arranged according to the Japanese list of causes of death and grouped in quinquennial age intervals formed the basis for the construction of the final life table which was completed in November 1919. This table, which like Mr. Fisher's other tables was derived without any information of the number of

lives exposed to risk at various ages, is shown in the addenda of this treatise.

Immediately after its construction Mr. Fisher sent this table to the well known Japanese actuary, Mr. T. Yano, and asked him for an opinion regarding the trustworthiness of the final death rates of q_x as derived by his new method. The Japanese actuary's answer arrived in April 1920. Mr. Yano had after the receipt of Mr. Fisher's letter ascertained the exposures and deaths among male lives at each seperate age for about 40 Japanese life offices during the period 1914—1917 and constructed by means of the conventional methods a complete series of q_x by integral ages from age 10 to 90. These ungraduated data are shown as a broken line polygon in the appended diagram (Figure 1). In spite of the fact that Mr. Fisher had no information whatever about the exposed to risk the agreement of the continuous curve of q_x as determined by the frequency curve method with Mr. Yano's ungraduated data is so close that I think further comments superfluous. The slight differences in younger ages might indeed rise from the fact that Mr. Yano had access to all the experience (containing more than 45,000 deaths) of all the Japenese companies, whereas Fisher only used the mortuary records as published by some of the leading Japanese companies.

Like all scientific methods of induction Mr. Fisher's proposed plan rests upon a working hypothesis, namely that it is possible from *biological* considerations to group the deaths among the survivors at various ages in any mortality table according to causes in such a manner that their percentage or relative frequency distribution according to attained

age at death will conform to a previously selected system or family of Laplacean-Charlier or Poisson-

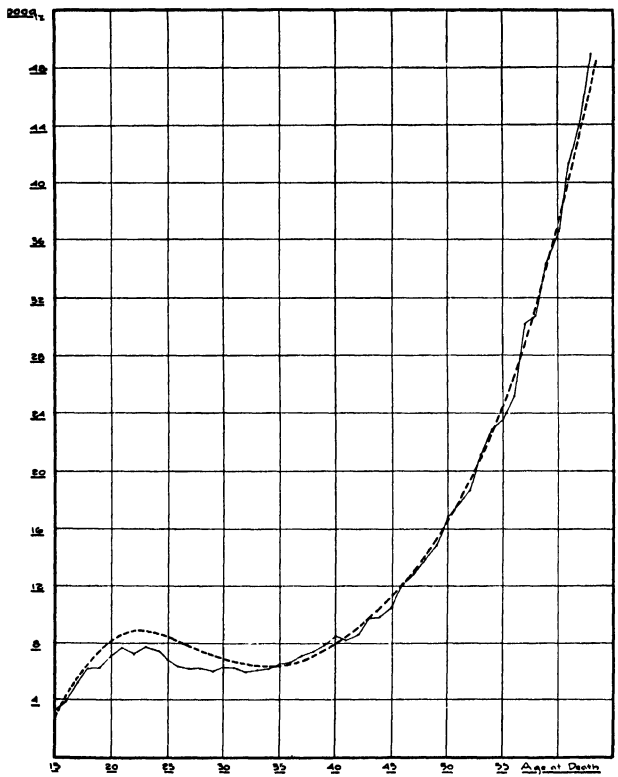

Fig. 1.

Charlier frequency curves. Mr. Fisher himself is very frank in stating that this is a working hypothesis

upon which hinges the success of the whole method. One of the main objections of his critics is that it seems impossible to prove the truth of this working hypothesis. Naturally its truth cannot be proved by mathematics or logic any more that we can prove or disprove the existence of Euclidean space, which in itself constitutes a working hypothesis for most of our applied mathematics. Mr. Fisher's critics might as well be asked to prove or disprove Newton's hypothetical laws of motion and attraction as extended by Maxwell and Hertz, or the newer hypothesis recently put forwards by the relativists, or the Lorentz hypothesis of contraction. It would indeed be a terriffic blow to science and the extension af knowledge if it was required that no working hypothesis would be alloved in scientific work unless such hypothesis could be proved to be true. What position would biology occupy to-day if biologists had insisted that Darwin's great hypothesis be proved before it could be allowed as a foundation in the study of evolution?

The most convincing answer to Mr. Fisher's captious critics among the old school of actuaries and statisticians is, however, the undisputed fact that his working hypothesis as such really does work. As pointed out by Dr. Pearl in the introductory note of this book the results set forth in the present treatise abundantly demonstrate this fact. The 6 widely different mortality tables as shown in the addenda stand as mute and yet as the most eloquent evidence to the fact that the method works. It might indeed not appear impertinent to suggest that Mr. Fisher's actuarial critics would render a greater service to their profession by proving that these six

mortality tables cannot be considered as reasonable approximatons to tables derived by orthodox means from the same population groups than by starting to poohpooh and ridicule his proposed method.

Winnipeg, Canada, November 1921.

E. A. Vigfusson.

"Nothing is less warranted in science than an uninquiring and unhoping spirit. In matters of this kind, those who despair are almost invariably those who have never tried to succeed."

W. Stanley Jevons.

CHAPTER I

(TRANSLATED BY MISS DICKSON)

AN INTRODUCTION TO THE THEORY OF FREQUENCY CURVES

1. INTRODUCTION The following method of constructing mortality tables from mortuary records by sex, age and cause of death rests essentially upon the theory of frequency curves originally introduced by the great Laplace and of recent years further developed and extended through the elegant and far reaching researches of the Scandinavian school of statisticians under the leadership of Gram, Charlier and Thiele and their disciples. This method is, however, comparatively little known and unfortunately not always fully appreciated by the majority of English statisticians and actuaries, who prefer to apply the well known methods of the eminent English biometrician, Karl Pearson. For this reason it may be advisable to give a preliminary sketch of Charlier's methods so as to obtain a better understanding of the

following chapters dealing with the more specific problem of mortality tables. The treatment must necessarily be brief and represents essentially an outline of the more detailed theory which I hope to present in my forthcoming second volume of the *Mathematical Theory of Probabilities*.

By the method of Charlier any frequency function is expressed as an infinite series rather than as a closed and compact algebraic or transcendental expression by the Pearsonian methods. By power series the thoughts of the majority of students are associated with the famous series which bear the names of Taylor and Maclaurin. In these series the function is derived as an infinite series of ascending powers of the independent variable whose coefficients are expressed by means of the correlated successive derivatives of the function for specific values of $f(x)$. Thus for instance we know that the Maclaurin series may be written as follows:

$$f(x) = f(0) + \frac{x}{\underline{1}} f'(0) + \frac{x^2}{\underline{2}} f''(0) + \ldots \frac{x^n}{\underline{n}} f^n(0) + \ldots$$

where $f^n(0)$ is the symbol for the value of the n^{th} derivative when $x = 0$ and $n = 1, 2, 3, 4 \ldots . n$.

There are, however, contrary to the belief of many immature students, only comparatively few functions which allow a rigorous expansion by this

method, in which the derived functions and the differential calculus play the leading roles.

But on the other hand there are other methods of expansions in infinite series which are more general and by which the coefficients of the independent variable are expressed by operations other than those of differentiation. One of these methods is to express the coefficients as definite integrals either of the unknown function itself or some auxiliary function.

The range of practical problems which lay themselves open to a successful attack along those lines is much wider than the corresponding range of practical problems to which we may apply the Taylor series.

Speaking generally as a layman (who continuously has to face practical rather than abstract problems) and specifically as a mathematical novice (who considers mathematics as a means rather than as an end) this fact appears to me quite obvious from a purely philosophical point of view. In nature and in all practical observations we encounter finite and not infinitesimal quantities. In other words, what we actually observe are finite sums or definite integrals, i. e. the limit of a sum of infinitely small component parts.

The definite integral rather than the derivative and the differential seems, therefore, to be the

more elementary and primitive operation and the one which suggests itself first hand. History of Mathematics indeed proves this contention. Archimedes had (as shown by the researches of the Danish scholar, Heiberg) laid the essential foundation for an integral calculus about 500 B. C. And nearly 25 centuries later, almost simultaneously with the historical discovery of Heiberg another Scandinavian, the Swedish mathematician and actuary, Fredholm, gave to the world his epochmaking work on integral equations. Fredholm's monumental memoir "*Sur une nouvelle methode pour la resolution du problems de Dirichlet*" was first published in the "*Öfversigt af akademiens forhandlinglar*" (Stockholm 1900). Measured by time the subject of integral equations is thus a mere infant in the history of mathematical discoveries. Measured by its importance it has already become a classic. Its application to a steadily increasing number of essentially practical problems in almost every branch of science has placed it in a central position of modern mathematical research and it bids fair to become the most important branch of mathematics.

Fredholm in introducing his now famous infinite determinants, known as the Fredholmean determinants, had a forerunner in the Danish actuary, Gram, whose Doctor's dissertation "Om

Rækkeudviklinger ved de mindste Kvadraters Metode" (Copenhagen 1879) gave prominence to a certain class of functions which later on have become known as orthogonal functions, and by which Gram actually gave the first expansion of a frequency distribution or frequency curve in an infinite series. Scandinavians in general and Scandinavian actuaries in particular may, therefore, feel proud of their share of imparting knowledge on this important subject, which makes a strong bid to place mathematics on a higher plane than ever before, not alone as an abstract but equally well as an applied science. The genius of the Italian renaissance Leonardo da Vinci, as early as 1479 proclaimed "that no part of human knowledge could lay claim to the title of science before it had passed through the stage of mathematical demonstration". Comparatively few branches of learning measure up to the standard of Leonardo da Vinci, and our learned friends among the economists and sociologists have a long road to travel before they succeed in placing their methods in the coveted niche of science. But the new vistas of possibilities opened up to them by means of M. Fredholm's discovery ought to furnish them a powerful tool towards the attainment of the high standard set by the great Italian.

The principal theorems of integral equations

are bound to be especially fruitful in their application to mathematical statistics and the problems of frequency curves and frequency surfaces together with the associated problems of mathematical correlation.

2. *FREQUENCY DISTRIBUTIONS AND FUNCTIONS* If N successive observations originating from the same essential circumstances or the same source of causes are made in respect to a certain statistical variate, x, and if the individual observations o_i $(i=1, 2, 3, \ldots . N)$ are permuted in an ascending order then this particular permutation is said to form a frequency distribution of x and is denoted by the symbol $F(x)$.

The relative frequencies of this specific permutation, that is the ratio which each absolute frequency or group of frequencies bear to the total number of observations, is called a relative frequency function or probability function and is denoted by the symbol $\varphi(x)$.

If the statistical variate is continuous or a graduated variate, such as heights of soldiers, ages at death of assured lives, physical and astronomical precision measurements, etc., then

$$dz\varphi(z)$$

is the probability that the variate x satisfies the following relation

$$z - \frac{1}{2}dz < x < z + \frac{1}{2}dz$$

or that x falls between the above limits.

If the statistical variate assumes integral (discrete) values only such as the number of alpha particles radiated from certain metals and radioactive gases as polonium and helium, number of fin rays in fishes, or number of petal flowers in plants, then $\varphi(z)$ is the probability that x assumes the value z. From the above definitions it follows a fortiori that

(a) $F(z) = N\varphi(z)$ (Integral variates)

(b) $dz\,F(z) = N\varphi(z)dz$ (Integrated variates)

Interpreting the above results graphically we find that (a) will be represented by a series of disconnected or discrete points while (b) will be represented by a continuous curve.

As to the function $\varphi(z)$ we make for the present no other assumptions than those following immediately from the customary definition of a mathematical probability. That is to say the function $\varphi(z)$ must be real and positive.

Moreover it must also satisfy the relation

$$\int_{-\infty}^{+\infty} \varphi(z)dz = 1,$$

or in the case of discrete variates:

$$\sum_{z=-\infty}^{z=\infty} \varphi(z) = 1$$

which is but the mathematical way of expressing the simple hypothetical disjunctive judgment that the variate is sure to assume some one or several values in the interval from $-\infty$ to $+\infty$. The zero point is arbitrarily chosen and need not coincide with the natural zero of the number scale. Thus for instance if we in the case of height of recruits choose the zero point of the frequency curve at 170 centimeters an observation of 180 centimeters would be recorded as $+10$ and an observation of 160 centimeters as -10.

3. PROPERTY OF CONSTANTS OR PARAMETERS In regard to a frequency function we may assume a priori that it will depend only upon the variate x and certain mathematical relations into which this variate enters with a number of constants $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \ldots\ldots$, symbolically expressed by the notation

$$F(x, \lambda_1, \lambda_2, \lambda_3, \lambda_4 \ldots .)$$

where the λ's are the constants and x the variate.

All these constants or parameters are naturally independent of x and represent some peculiar properties or characteristic essentials of the frequency

function as expressed in the original observations o_i ($i=1, 2, 3, \ldots\ldots N$). We may, therefore, say that each constant or statistical parameter entering into the final mathematical form for the frequency function is a function of the observations o_i. This fact may be expressed in the following symbolic form :—

$$\lambda_1 = S_1 (o_1, o_2, o_3, \ldots o_N)$$
$$\lambda_2 = S_2 (o_1, o_2, o_3, \ldots o_N)$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$\lambda_N = S_N (o_1, o_2, o_3, \ldots o_N).$$

But from purely a priori considerations we are able to tell something else about the function S_i ($i=1, 2, 3 \ldots N$). It is only when permuting the various o's in an ascending magnitude according to the natural number scale that we obtain a frequency function. This arrangement itself has, however, no influence upon any one of the o's which were generated before this purely arbitrary permutation took place. The ultimate and previously measured effects of the causes as reflected in each individual numerical observations, o_i, depend only upon the origin of causes which form the fundamental basis for the statistical object under investigation and do not depend

upon the order in which the individual o's occur in the series of observations.

Suppose for instance that the observations occurred in the following order

$$o_1,\ o_2,\ o_3,\ \ldots\ldots\ o_N.$$

By permuting these elements in their natural order we obtain the frequency distribution $F(x)$. But the very same distribution could have been obtained if the observations had occurred in any other order as for instance

$$o_7,\ o_9,\ o_N\ ,\ \ldots\ o_3\ \ldots\ o_1.$$

so long as all of the individual o's were retained in the original records. Or to take a concrete example as the study of the number of policyholders according to attained ages in a life assurance office. We write the age of each individual policyholder on a small card. When all the ages have been written on individual cards they may be permuted according to attained age and the resulting series is a frequency function of the age x. We may now mix these cards just as we mix ordinary playing cards in a game of whist, and we get another permutation—in general different from the order in which we originally recorded the ages on the cards. But this new permutation can equally

well be used to produce the frequency function if we are only sure to retain all the cards and do not add any new cards.

4. PARAMETERS-VIEWED AS SYMMETRIC FUNCTIONS

The various functions $S(o_1, o_2, o_3 \ldots\ldots o_N)$ are therefore, symmetric functions, that is functions which are left unaltered by arbitrarily permuting the N elements o, and no interchange whatever of the values of the various o's in those symmetric functions can have any influence upon the final form of the frequency function or frequency curve, $F(x)$.

We now introduce under the name of *power sums* a certain well known form of fundamental symmetrical functions defined by the following relations

$$s_0 = o_1^0 + o_2^0 + o_3^0 + \ldots o_N^0 = N$$
$$s_1 = o_1^1 + o_2^1 + o_3^1 + \ldots o_N^1 = \sum o_i^1$$
$$s_2 = o_1^2 + o_2^2 + o_3^2 + \ldots o_N^2 = \sum o_i^2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$s_N = o_1^N + o_2^N + o_3^N + \ldots o_N^N = \sum o_i^N.$$

Moreover, a well known theorem in elementary algebra tells us that every symmetric function may be expressed as a function of $s_1, s_2, s_3 \ldots \ldots s_N$.

From this theorem it follows a fortiori that we are able to express the constants λ in the frequency curve as functions of the power sums of the observations. While such a procedure is possible, theoretically at least, we should, however, in most cases find it a very tedious and laborious task in actual practice. It, therefore, remains to be seen whether it is possible to transform these symmetrical functions of the power sums of the observations into some other symmetric functions, which are more flexible and workable in practical computations and which can be expressed in terms of the various values of s.

5. THIELE'S SEMI-INVARIANTS

It is the great achievement of Thiele to have been the first mathematician to realize this possibility and make this transformation by introducing into the theory of frequency curves a peculiar system of symmetrical functions which he called *semi invariants* and denoted by the symbols $\lambda_1, \lambda_2, \lambda_3 \ldots$

Starting with power sums, s_i. Thiele defines these by the following identity

$$s_0 e^{\frac{\lambda_1\omega}{\lfloor 1} + \frac{\lambda_2\omega^2}{\lfloor 2} + \frac{\lambda_3\omega^3}{\lfloor 3} + \cdots} = s_0 + \frac{s_1\omega}{\lfloor 1} + \frac{s_2\omega^2}{\lfloor 2} + \frac{s_3\omega^3}{\lfloor 3} + \ldots \quad (1)$$

which is identical in respect to ω.

Since $s_i = \sum o^i$ the right hand side of the equation may also be written as $e^{o_1\omega} + e^{o_2\omega} + e^{o_3\omega} + \ldots =$
$= \sum e^{o_i\omega}$.

Differentiating (1) with respect to ω we have

$$s_0 e^{\frac{\lambda_1\omega}{|1} + \frac{\lambda_2\omega^2}{|2} + \frac{\lambda_3\omega^3}{|3} + \ldots} \left[\lambda_1 + \frac{\lambda_2\omega}{|1} + \frac{\lambda_3\omega^2}{|2} + \ldots\right] =$$

$$= \left[s_0 + \frac{s_1}{|1}\omega + \frac{s_2}{|2}\omega^2 + \frac{s_3}{|3}\omega^3 + \ldots\right] \left[\lambda_1 + \frac{\lambda_2\omega}{|1} + \frac{\lambda_3}{|2}\omega^2 + \ldots\right] =$$

$$= s_1 + \frac{s_2}{|1}\omega + \frac{s_3}{|2}\omega^2 + \frac{s_4}{|3}\omega^3 + \ldots$$

Multiplying out and equating the various coefficients of equal powers of ω we finally have

$$\begin{aligned}
s_1 &= \lambda_1 s_0 \\
s_2 &= \lambda_1 s_1 + \lambda_2 s_0 \\
s_3 &= \lambda_1 s_2 + 2\lambda_2 s_1 + \lambda_3 s_0 \\
s_4 &= \lambda_1 s_3 + 3\lambda_2 s_2 + 3\lambda_3 s_1 + \lambda_4 s_0 \\
&\ldots\ldots\ldots\ldots\ldots \\
&\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

where the coefficients follow the law of the binomial theorem.

Solving for λ we have

$$\begin{aligned}
\lambda_1 &= s_1 : s_0 \\
\lambda_2 &= (s_2 s_0 - s_1^2) : s_0^2 \\
\lambda_3 &= (s_3 s_0^2 - 3 s_2 s_1 s_0 + 2 s_1^3) : s_0^3
\end{aligned}$$

$$\lambda_4 = s_4 s_0^3 - 4 s_3 s_1 s_0^2 - 3 s_2^2 s_0^2 + 12 s_2 s_1^2 s_0 - 6 s_1^4) : s_0^4$$

. . . . . .

. . . . . .

The semi-invariants λ in respect to an arbitrary origin and unit are as we noted defined by the relation

$$s_0 e^{\frac{\lambda_1 \omega}{\lfloor 1} + \frac{\lambda_2 \omega^2}{\lfloor 2} + \frac{\lambda_3 \omega^3}{\lfloor 3} + \cdots} = e^{o_1 \omega} + e^{o_2 \omega} + e^{o_3 \omega} + \ldots$$

where o_1, o_2, o_3 . . . are the individual observations.

Let us now change to another coordinate system with another unit and origin defined by the following linear transformations:—

$$o_i' = a o_i + c \qquad (i = 1, 2, 3, \ldots).$$

The semi-invariants in this new system are given by the relation

$$s_0 e^{\frac{\lambda_1' \omega}{\lfloor 1} + \frac{\lambda_2' \omega^2}{\lfloor 2} + \frac{\lambda_3' \omega^3}{\lfloor 3} + \cdots} = e^{o_1' \omega} + e^{o_2' \omega} + e^{o_3' \omega} + \ldots =$$

$$= e^{(a o_1 + c)\omega} + e^{(a o_2 + c)\omega} + \ldots$$

Since the various values of λ' do not depend upon the quantity ω we may without changing the value of the semi-invariants replace ω by $\omega : a$ in the above equations, which gives

$$s_0 e^{\frac{\lambda'_1\omega}{\lfloor 1\, a}+\frac{\lambda'_2\omega^2}{a^2\lfloor 2}+\frac{\lambda'_3\omega^3}{a^3\lfloor 3}+\cdots} =$$

$$= e^{(ao_1+c)\frac{\omega}{a}} + e^{(ao_2+c)\frac{\omega}{a}} + e^{(ao_3+c)\frac{\omega}{a}} + \ldots =$$

$$= e^{\frac{c\omega}{a}}\left[e^{o_1\omega}+e^{o_2\omega}+e^{o_3\omega}+\ldots\right] =$$

$$= e^{\frac{c\omega}{a}}\, s_0 e^{\frac{\lambda_1\omega}{\lfloor 1}+\frac{\lambda_2\omega^2}{\lfloor 2}+\frac{\lambda_3\omega^3}{\lfloor 3}+\cdots}$$

Taking the logarithms on both sides of the equation we have

$$\frac{\lambda'_1\omega}{a\lfloor 1}+\frac{\lambda'_2\omega^2}{a^2\lfloor 2}+\frac{\lambda'_3\omega^3}{a^3\lfloor 3}+\ldots =$$

$$= \frac{c\omega}{a}+\frac{\lambda_1\omega}{\lfloor 1}+\frac{\lambda_2\omega^2}{\lfloor 2}+\frac{\lambda_3\omega^3}{\lfloor 3}+\ldots$$

Differentiating successively with respect to ω we have

$$\frac{\lambda'_1}{a\lfloor 1}+\frac{\lambda'_2\omega}{a^2}+\frac{\lambda'_3\omega^3}{2a^3}+\ldots = \frac{c}{a}+\lambda_1+\lambda_2\omega+\frac{\lambda_3\omega^2}{2}+\ldots$$

$$\frac{\lambda'_2}{a^2}+\frac{\lambda'_3\omega}{a^3}+\frac{\lambda'_4\omega^2}{2a^4}+\ldots = \lambda_2+\lambda_3\omega+\frac{\lambda_4\omega^2}{\lfloor 2}+\ldots$$

$$\frac{\lambda'_3}{a^3}+\frac{\lambda'_4\omega}{a^4}+\ldots = \lambda_3+\lambda_4\omega+\ldots$$

Letting $\omega = 0$ we therefore have

$$\frac{\lambda'_1}{a} = \frac{c}{a} + \lambda_1 \text{ or } \lambda'_1 = a\lambda_1 + c$$

$$\frac{\lambda'_2}{a^2} = \lambda_2 \text{ or } \lambda'_2 = a^2\lambda_2$$

$$\frac{\lambda'_3}{a^3} = \lambda_3 \text{ or } \lambda'_3 = a^3\lambda_3$$

. . . . . .

. . . . . .

from which we deduce the following relations

$$\lambda_1(ax + c) = a\lambda_1(x) + c$$
$$\lambda_r(ax + c) = a^r\lambda_r(x) \quad \text{for } r > 1,$$

which shows how the semi-invariants change by introducing a new origin and a new unit.

We shall for the present leave the semi invariants and only ask the reader to bear in mind the above relations between λ and s, of which we shall later on make use in determining the constants in the frequency curve $\varphi(x)$.

6. *THE FOURIER INTEGRALS* Before discussing the generation of the total frequency curve it will, however, be necessary to demonstrate some auxiliary mathematical formulae from the theory of definite integrals and integral equations which will be of use in the

following discussion as mathematical tools with which to attack the collected statistical data or the numerical observations.

One of these tools is found in the celebrated integral theorem by Fourier, which was the first integral equation to be successfully treated. We shall in the following demonstration adhere to the elegant and simple solution by M. Charlier. Charlier in his proof supposes that a function, $F(\omega)$, is defined through the following convergent series.

$$F(\omega) = \alpha [f(o) + f(\alpha) e^{\alpha\omega i} + f(2\alpha) e^{2\alpha\omega i} + \dots$$

$$+ f(\alpha) e^{-\alpha\omega i} + f(-2\alpha) e^{-2\alpha\omega i} + \dots$$

or
$$F(\omega) = \alpha \sum_{m=-\infty}^{m=\infty} f(\alpha m) e^{\alpha m \omega i} \qquad (2)$$

where $i = \sqrt{-1}$.

We then see by the well known theorem of Cauchy that the integral

$$I(\omega) = \int_{-\infty}^{+\infty} f(x) e^{x\omega i} dx \qquad (3)$$

is finite and convergent. If we now let $m\alpha = x$ and let $\alpha = 0$ as a limiting value, α becomes equal to dx and $f(\alpha m) = f(x)$. Consequently we may write

$$\lim_{\alpha = 0} F(\omega) = I(\omega).$$

Multiplying (2) by $e^{-r\alpha\omega i}\,d\omega$ and integrating between the limits $-\pi/\alpha$ and $+\pi/\alpha$ we get on the left an expression of the form

$$\int_{-\pi/\alpha}^{+\pi/\alpha} F(\omega)\, e^{-r\alpha\omega i}\, d\omega$$

and on the right a sum of definite integrals of which, however, all but the term containing $f(r\alpha)$ as a factor will vanish. This particular term reduces to

$$\alpha \int_{-\pi/\alpha}^{+\pi/\alpha} f(r\alpha)\, d\omega \quad \text{or} \quad 2\pi f(r\alpha).$$

Hence we have

$$f(r\alpha) = \frac{1}{2\pi} \int_{-\pi/\alpha}^{+\pi/\alpha} F(\omega)\, e^{-r\alpha\omega i}\, d\omega. \qquad (4a)$$

By letting α converge toward zero and by the substitution $r\alpha = x$ this equation reduces to

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} I(\omega)\, e^{-x\omega i}\, d\omega. \qquad (4b)$$

Charlier has suggested the name *conjugated Fourier function* of $f(x)$ for the expression $F(\omega)$. We then have, if we introduce a new function $\psi(\omega)$ defined by the simple relation:

$$\sqrt{2\pi}\,\psi(\omega) = \lim_{\alpha = 0} F(\omega)$$

$$\psi(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{x\omega i}\, dx. \qquad (5\,a)$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(\omega) e^{-x\omega i}\, d\omega. \qquad (5\,b)$$

The equations (5a) and (5b) are known as integral equations of the first kind. The expression $e^{x\omega i}$ (or $e^{-x\omega i}$) is known as the *nucleus* of the equation. If in (5b) we know the value of $\psi(\omega)$ we are able to determine $f(x)$. Inversely, if we know $f(x)$ we may find $\psi(\omega)$ from (5a).

7. *FREQUENCY CURVE AS THE SOLUTION OF AN INTEGRAL EQUATION*

We are now in a position to make use of the semi-invariants of Thiele, which hitherto in our discussion have appeared as a rather disconnected and alien member. On page 13 we saw that the semi-invariants could be expressed by the relation

$$e^{\frac{\lambda_1}{\lfloor 1}\omega + \frac{\lambda_2}{\lfloor 2}\omega^2 + \frac{\lambda_3}{\lfloor 3}\omega^3 + \cdots} = \sum e^{o_i\omega}$$

where o_i $(i = 1, 2, 3 \ldots\ldots)$ denotes the individual observations.

The definition of the *semi-invariants* does not necessitate that all the o's must be different. If some of the o's are exactly alike it is self-evident that the term $e^{o_i\omega}$ must be repeated as often as o occurs among all of the observations. If therefore $N\varphi(o_i)$ denotes the absolute frequency of o_i where $\varphi(o_i)$ is the relative frequency function, then the definition of the semi-invariants may be written as:—

$$\sum \varphi(o_i) e^{\frac{\lambda_1}{\lfloor 1}\omega + \frac{\lambda_2}{\lfloor 2}\omega^2 + \frac{\lambda_3}{\lfloor 3}\omega^3 + \cdots} = \sum \varphi(o_i) e^{o_i\omega}.$$

For continuous variates, x, the above sums are transformed into definite integrals of the form

$$e^{\frac{\lambda_1}{\lfloor 1}\omega + \frac{\lambda_2}{\lfloor 2}\omega^2 + \frac{\lambda_3}{\lfloor 3}\omega^3 + \cdots} \int_{-\infty}^{+\infty} \varphi(x)\,da = \int_{-\infty}^{+\infty} \varphi(x) e^{x\omega}\,dx.$$

Let us now substitute the quantity $\omega\sqrt{-1}$, or $i\omega$, for ω in the above identity. We then have:—

$$e^{\frac{\lambda_1}{\lfloor 1}i\omega + \frac{\lambda_2}{\lfloor 2}i^2\omega^2 + \frac{\lambda_3}{\lfloor 3}i^3\omega^3 + \cdots} \int_{-\infty}^{+\infty} \varphi(x)\,dx = \int_{-\infty}^{+\infty} \varphi(x) e^{ix\omega}\,dx$$

under the supposition that this transformation holds in the complex region in which the function is defined.

In this equation the definite integrals are of special importance. The factor $\int_{-\infty}^{+\infty} \varphi(x)\,dx$ is, of course, equal to unity according to the simple considerations set forth on page seven. The integral on the right hand side of the equation is, however, apart from the constant factor $\sqrt{2\pi}$ nothing more than the ψ function in the conjugate Fourier function if we let $\varphi(x) = f(x)$, and

$$e^{\frac{\lambda_1}{\lfloor 1}i\omega + \frac{\lambda_2}{\lfloor 2}i^2\omega^2 + \frac{\lambda_3}{\lfloor 3}i^3\omega^3 + \cdots} = \sqrt{2\pi}\,\psi(\omega).$$

According to (5b) we may, therefore write $f(x)$ or $\varphi(x)$ as

$$\varphi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{\frac{\lambda_1}{\lfloor 1}i\omega + \frac{\lambda_2}{\lfloor 2}i^2\omega^2 + \frac{\lambda_3}{\lfloor 3}i^3\omega^3 + \cdots} e^{-x\omega i}\,d\omega$$

as the most general form of the frequency function $\varphi(x)$ expressed by means of semi-invariants.

8. FIRST APPROXIMATE SOLUTION The exactness with which $\varphi(x)$ is reproduced depends, of course, upon the number of λ's we decide to consider in the above formula. As a first approximation we may omit all λ's

above the order 2 or all terms in the exponent with indices higher than 2. Bearing in mind that $i^2 = -1$ we therefore have as a first approximation

$$\varphi_0(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega(\lambda_1 - x) - \frac{\lambda_2}{2}\omega_2} d\omega.$$

The above definite integral was first evaluated by Laplace by means of the following elegant analysis. Using the well known Eulerean relation for complex quantities the above integral may be written as

$$\int_{-\infty}^{+\infty} e^{-\frac{\lambda_2 \omega^2}{2}} \cos\left[(\lambda_1 - x)\omega\right] d\omega +$$

$$+ i \int_{\infty}^{+\infty} e^{-\frac{\lambda^2}{2}\omega^2} \sin\left[(\lambda_1 - x)\omega\right] d\omega.$$

The imaginary member vanishes because the factor $e^{-\frac{\lambda_2}{2}\omega^2}$ is an even function and $\sin\left[(\lambda_1 - x)\omega\right]$ an uneven function, the area from $-\infty$ to 0 will therefore equal the area from 0 to $+\infty$, but be opposite in sign, which reduces the total area from $-\infty$ to $+\infty$ or the integral in question to zero.

In regard to the first term, similar conditions hold except that $\cos\left[(\lambda_1 - x)\omega\right]$ is an even function and the integral may hence be written as

$$I = 2\int_0^\infty e^{-\frac{\lambda_2}{2}\omega^2}\cos(r\omega)\,d\omega \quad \text{where } r = \lambda_1 - x.$$

Regarding the parameter r as a variable and differentiating I in respect to this variable we have

$$\frac{dI}{dr} = \frac{2}{\lambda_2}\int_0^\infty \left(-\lambda_2\omega e^{-\frac{\lambda_2}{2}\omega^2}\right)\sin(r\omega)\,d\omega.$$

From this we have by partial integration :—

$$\frac{dI}{dr} = \frac{2}{\lambda_2}\left[e^{-\frac{\lambda_2}{2}\omega^2}\sin(r\omega)\,d\omega\right]_0^\infty - \frac{2r}{\lambda_2}\int_0^\infty e^{-\frac{\lambda_3}{2}\omega^2}\cos(r\omega)\,d\omega$$

$$= 0 - \frac{rI}{\lambda_2} \quad \text{or} \quad \frac{1}{I}\frac{dI}{dr} = -\frac{r}{\lambda_2}.$$

From which we find

$$\log I = \frac{-r^2}{2\lambda} + \log A$$

where $\log A$ is a constant. Hence we have :—

$$I = Ae^{-\frac{r^2}{2\lambda_2}}.$$

In order to determine A we let $r = 0$ and we have

$$I_0 = A = 2\int_0^\infty e^{-\frac{\lambda_2}{2}\omega_2} d\omega = 2\sqrt{\frac{\pi}{2\lambda_2}} = \sqrt{\frac{2\pi}{\lambda_2}}$$

This finally gives the expression for $\varphi_0(x)$ in the following form:

$$\varphi_0(x) = \frac{1}{\sqrt{2\pi\lambda_2}} e^{-\frac{(\lambda_1 - x)^2}{2\lambda_2}}$$

as a preliminary approximation for the frequency curve $\varphi(x)$.

The first mathematical deduction of this approximate expression for a frequency curve is found in the monumental work by Laplace on Probabilities, and the function $\varphi_0(x)$ entering in the expression $\varphi_0(x)\ dx$, which gives the probability that the variate will fall between $x - \frac{1}{2}dx$ and $x + \frac{1}{2}dx$, is therefore known as the Laplacean probability function or sometimes as the Normal Frequency Curve of Laplace. The same curve was, as we have mentioned also previously deduced independently by Gauss in connection with his studies on the distribution of accidental errors in precision measurements.

Laplace's probability function, $\varphi_0(x)$ possesses some remarkable properties which it might

be well worth while to consider. Introducing a slightly different system of notation by writing $\lambda_1 = M$ and $\sqrt{\lambda_2} = \sigma$, $\varphi_0(x)$ reduces to the following form.

$$\varphi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-M)^2 : 2\sigma^2}$$

which is the form introduced by Pearson.

The frequency curve, $\varphi_0(x)$, is here expressed in reference to a Cartesian coordinate system with origin at the zero point of the natural number system and whose unit of measurement is also equivalent to the natural number unit. It is, however, not necessary to use this system in preference to any other system. In fact, we may choose arbitrarily any other origin and any other unit standard without altering the properties of the curve. Suppose, therefore, that we take M as the origin and σ as the unit of the system. The frequency function then reduces to

$$\varphi_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2 : 2}.$$

Since the integral of $\varphi_0(x)$ from $-\infty$ to $+\infty$ equals unity the following equation must necessarily hold.

$$\int_{-\infty}^{+\infty} e^{x^2 : 2} = \sqrt{2\pi}.$$

9. DEVELOPMENT BY POLYNOMIALS The Laplacean Probability Curve possesses, however, some other remarkable properties which are of great use in expanding a function in a series. Starting with $\varphi_0(x)$ we may by repeated differentiation obtain its various derivaties. Denoting such derivatives by $\varphi_1(x)$, $\varphi_2(x)$, $\varphi_3(x)$ . . . respectively we have the following relations.[1])

$$\begin{aligned}
\varphi_0(x) &= e^{-x^2:2} \\
\varphi_1(x) &= -x\varphi_0(x) \\
\varphi_2(x) &= (x^2-1)\varphi_0(x) \\
\varphi_3(x) &= -(x^3-3x)\varphi_0(x) \\
\varphi_4(x) &= (x^4-6x^2+3)\varphi_0(x) \\
&\cdots\cdots\cdots\cdots\cdots \\
&\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

and in general for the n^{th} derivative:—

$$\varphi_n(x) = (-1)^n\left[x^n - \frac{n(n-1)}{\lfloor 2}x^{n-2} + \right.$$

$$+\frac{n(n-1)(n-2)(n-3)}{2\cdot 4}x^{n-4}$$

$$\left.-\frac{n(n-1)(n-2)(n-3)(n-4)(n-5)}{2\cdot 4\cdot 6}x^{n-6}+\ldots\right]\varphi_0(x).$$

[1] In the following computations we have omitted temporarily the constant factor $1:\sqrt{2\pi}$ of $\varphi_0(x)$ and its derivatives.

It can be readily seen that the derivatives of $\varphi_0(x)$ are represented throughout as products of polynomials of x and the function $\varphi_0(x)$ itself. The various polynomials

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= -x \\ H_2(x) &= x^2 - 1 \\ H_3(x) &= -(x^3 - 3x) \\ H_4(x) &= (x^4 - 6x^2 + 3) \end{aligned}$$

and so forth are generally known as Hermite's polynomials from the name of the French mathematician, Hermite, who first introduced these polynomials in mathematical analysis.

The following relations can be shown to exist between the three polynomials

$$H_{n+1}(x) - xH_n(x) + nH_{n-1}(x) = 0$$

and

$$\frac{d^2H_n(x)}{dx^2} - \frac{xdH_n(x)}{dx} + nH_n(x) = 0.$$

A numerical 10 decimal place tabulation of the first six Hermite polynomials for values of x up to 4 and progressing by intervals of 0.01 is given by Jørgensen in his Danish work "Frekvensflader og Korrelation".

There exist now some very important relations between the Hermite polynomials and the derivatives of $\varphi_0(x)$, or between $H_n(x)$ and $\varphi_n(x)$.

Consider for the moment the two following series of functions

$$\varphi_0(x),\ \varphi_1(x),\ \varphi_2(x),\ \varphi_3(x),\ \varphi_4(x),\ \ldots$$
$$H_0(x),\ H_1(x),\ H_2(x),\ H_3(x),\ H_4(x),\ \ldots$$

where $\varphi_n(x) = H_n(x)\varphi_0(x)$ and where $\lim \varphi_n(x) = 0$ for $x = \pm\infty$.

We shall now prove that the two series $\varphi_n(x)$ and $H_n(x)$ form a biorthogonal system in the interval $-\infty$ to $+\infty$, that is to say that they are

(1) real and continuous in the whole plane

(2) no one of them is identically zero in the plane

(3) every pair of them $\varphi_n(x)$ and $H_m(x)$, satisfy the relation.

$$\int_{-\infty}^{+\infty} \varphi_n(x) H_m(x)\, dx = 0 \qquad (n \gtrless m).$$

We have the self evident relation (letting $x = z$)

$$\int_{-\infty}^{+\infty} H_m(z)\varphi_n(z)\, dz = \int_{-\infty}^{+\infty} H_m(z) H_n(z)\varphi_0(z)\, dz =$$
$$= \int_{-\infty}^{+\infty} H_n(z)\varphi_m(z)\, dz.$$

Since this relation holds for all values of m and n it is only necessary to prove the proposition for $n > m$. For if it holds for $n > m$ it will according to the above relation also hold for $n < m$.

By partial integration we have :—

$$\int_{-\infty}^{+\infty} H_m(z)\,\varphi_n(z)\,dz =$$

$$= H_m(z)\,\varphi_{n-1}(z)\Big]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} H'_m(z)\,\varphi_{n-1}(z)\,dz$$

when $H'_m(z)$ is the first derivative of $H_m(z)$.

The first member on the right reduces to 0 since $\varphi_{n-1}(z) = 0$ for $z = \pm\infty$. We have therefore :—

$$\int_{-\infty}^{+\infty} H_m(z)\,\varphi_n(z)\,dz = -\int_{-\infty}^{+\infty} H'_m(z)\,\varphi_{n-1}(z)\,dz$$

$$\int_{-\infty}^{+\infty} H'_m(z)\,\varphi_{n-1}(z)\,dz = -\int_{-\infty}^{+\infty} H''_m(z)\,\varphi_{n-2}(z)\,dz$$

$$\int_{-\infty}^{+\infty} H''_m(z)\,\varphi_{n-2}(z)\,dz = -\int_{-\infty}^{+\infty} H'''_m(z)\,\varphi_{n-3}(z)\,dz.$$

Continuing this process we obtain finally an expression of the form

$$\int_{-\infty}^{+\infty} H_m(z)\,\varphi_n(z)\,dz = (-1)^{m+1}\int_{-\infty}^{+\infty} H_m^{(m+1)}\varphi_{n-m-1}(z)\,dz,$$

when $H_m^{(m+1)}(z)$ is the $m+1$ derivative of $H_m(z)$ and $n-m-1>0$. Since $H_m(z)$ is a polynomial in the m^{th} degree its $m+1$ derivative is zero and we have finally that

$$\int_{-\infty}^{+\infty} H_m(z)\varphi_n(z)dz = 0$$

for all values of m and n where $\gtrless m$.

For $m = n$ we proceed in exactly the same manner, but stop at the m^{th} integration. We have, therefore, by replacing m by n in the above partial integrations

$$\int_{-\infty}^{+\infty} H_n(z)\varphi_n(z)dz = (-1)^n \int_{-\infty}^{+\infty} H_n^{(n)}(z)\varphi_{n-n}(z)dz =$$

$$= (-1)^n \int_{-\infty}^{+\infty} H_n^{(n)}(z)\varphi_0(z)dz.$$

The n^{th} derivative of $H_n(z)$ is, however, nothing but a constant and equal to $(-1)^n \lfloor n$. Hence we have finally

$$\int_{-\infty}^{+\infty} H_n(z)\varphi_n(z)dz = (-1)^n(-1)^n \lfloor n \int_{-\infty}^{+\infty} e^{-z^2:2}dz =$$

$$= \lfloor n \sqrt{2\pi}.$$

The above analysis thus proves that the functions $H_m(z)$ and $\varphi_n(z)$ are biorthogonal to each other for all values of n different from m throughout the whole plane.

We can now make use of these relations between the infinite set of biorthogonal functions $H_m(z)$ and $\varphi_n(z)$ in solving the problem of ex-

panding an arbitrary function $\varphi(z)$ in a series of the form

$$\varphi(z) = c_0\varphi_0(z) + c_1\varphi_1(z) + c_2\varphi_2(z) + \ldots$$

the series to hold in the interval from $-\infty$ to $+\infty$.

If we know that $\varphi(z)$ can be developed into a series of this form, which after multiplication by any continuous function can be integrated term for term, then we are are able to give a formal determination of the coefficients c.

This formal determination of any one of the c's, say c_i consists in multiplying the above series by $H_i(z)$ and integrating each term from $-\infty$ to ∞. All the terms except the one containing the product $H_i(z)\varphi_i$ vanish and we have for c_i.

$$c_i = \frac{\int_{-\infty}^{+\infty} \varphi(z)H_i(z)dz}{\int_{-\infty}^{+\infty} \varphi_i(z)H_i(z)dz} = \frac{\int_{-\infty}^{+\infty} \varphi(z)H_i(z)dz}{\lfloor i \sqrt{2\pi}} .$$

If we define the Hermite functions as

$$\begin{aligned} H_0(z) &= 1 \\ H_1(z) &= z \\ H_2(z) &= z^2 - 1 \\ H_3(z) &= z^3 - 3z \\ H_4(z) &= z^4 - 6z^2 + 3 \end{aligned}$$

the above formula takes on the form

$$c_i = \frac{\int_{-\infty}^{+\infty} \varphi(z)\, H_i(z)\, dz}{\int_{-\infty}^{+\infty} \varphi_i(z)\, H_i(z)\, dz} = \frac{\int_{-\infty}^{+\infty} \varphi(z)\, H_i(z)\, dz}{(-1)^r \lfloor i \sqrt{2\pi}}$$

which we shall prefer to use in the following discussion.

It will be noted that this purely formal calculation of the coefficients c is very similar to the determination of the constants in a Fourier Series, where as a matter of fact the system of functions

$$\cos z,\ \cos 2z,\ \cos 3z,\ \ldots\ldots$$
$$\sin z,\ \sin 2z,\ \sin 3z,\ \ldots\ldots$$

is biorthogonal in the interval $0 \leq z \leq 1$.

But the reader must not forget that the above representation is only a formal one, and we do not know if it is valid. To prove its validity we must first show that the series is convergent and secondly that it actually represents $\varphi(z)$ for all values of z.

This is by no means a simple task and it cannot be done by elementary methods. A Russian mathematician, Vera Myller-Lebedeff, has, however, given an elegant solution by means of some well known theorems from the Fredholm integral

equations. She has among other things proved the following criterion:—

"Every function $\varphi(z)$ which together with its first two derivatives is finite and continuous in the interval from $-\infty$ to $+\infty$ and which vanishes together with its derivatives for $z = \pm\infty$ can be developed into an infinite series of the form:—

$$\varphi(z) = \sum c_i e^{-z^2:2} H_i(z)$$

where $H_i(z)$ is the Hermite polynomial of order i".

10. GRAM'S SERIES It is, however, not our intention to follow up this treatment which is outside the scope of an elementary treatise like this and shall in its place give an approximate representation of the frequency function, $\varphi(z)$, by a method, which in many respects is similar to that introduced by the Danish actuary Gram in his epochmaking work "Udviklingsrækker", which contains the first known systematic development of a skew frequency function. Gram's problem in a somewhat modified form may briefly be stated as follows:—*Being given an arbitrary relative frequency function, $\varphi(z)$, continuous and finite in the interval $-\infty$ to $+\infty$ (and which vanishes*

for $z = \pm\infty$) to determine the constant coefficients c_0, c_1, c_2, c_3 . . . . . in such a way that the series

$$\frac{c_0\varphi_0(z)}{\sqrt{\varphi_0(z)}} + \frac{c_1\varphi_1(z)}{\sqrt{\varphi_0(z)}} + \frac{c_2\varphi_2(z)}{\sqrt{\varphi_0(z)}} + \ldots + \frac{c_n\varphi_n(z)}{\sqrt{\varphi_0(z)}} =$$

$$= \frac{1}{\sqrt{\varphi_0(z)}} \sum c_i\varphi_i(z)$$

gives the best approximation to the quantity $\varphi(z) : \sqrt{\varphi_0(z)}$ in the sense of the method of least squares. That is to say we wish to determine the constants c in such a manner that the sum of the squares of the differences between the function and the approximate series becomes a minimum. This means that the expression

$$I = \int_{-\infty}^{+\infty} \left[\frac{\varphi(z)}{\sqrt{\varphi_0(z)}} - \sum \frac{c_i\varphi_i(z)}{\sqrt{\varphi_0(z)}}\right]^2 dz$$

must be a minimum.

On the basis of this condition we have

$$\frac{\varphi(z)}{\sqrt{\varphi_0(z)}} \lessgtr \frac{1}{\sqrt{\varphi_0(z)}} \sum c_i\varphi_i(z) = \sqrt{\varphi_0(z)} \sum c_i H_i(z) = U(z)$$

where the unknown coefficients c must be so determined that

$$I = \int_{-\infty}^{+\infty} \left[\frac{\varphi(z)}{\sqrt{\varphi_0(z)}} - U(z) \right]^2 dz \qquad \text{equals a minimum.}$$

Taking the partial derivatives in respect to c_i we have

$$\frac{\delta I}{\delta c_i} = -\frac{2\delta}{\delta c_i} \int_{-\infty}^{+\infty} \frac{\varphi(z)}{\sqrt{\varphi_0(z)}} U(z)\, dz + \frac{\delta}{\delta c_i} \int_{-\infty}^{\infty} [U(z)]^2 dz.$$

Now since

$$\int_{\infty}^{\infty} [U(z)]^2 dz =$$

$$\int_{-\infty}^{+\infty} \left\{ c_0^2 [H_0(z)]^2 + c_1^2 [H_1(z)]^2 + \ldots c_n^2 [H_n(z)]^2 \right\} \varphi_0(z)\, dz,$$

we get

$$\frac{\delta I}{\delta c_i} = -2 \int_{-\infty}^{+\infty} \frac{\varphi(z)}{\sqrt{\varphi_0(z)}} H_i(z) \sqrt{\varphi_0(z)}\, dz + 2 c_i \int_{-\infty}^{+\infty} [H_i(z)]^2 \varphi_0(z)\, dz$$

where the latter integral equals

$$\int_{-\infty}^{+\infty} \varphi_i(z) H_i(z)\, dz = (-1)^i \lfloor i \sqrt{2\pi}.$$

Equating to zero and solving for c_i we finally obtain the following value for c_i—

$$c_i = \frac{(-1)^i}{\lfloor i \sqrt{2\pi}} \int_{-\infty}^{+\infty} \varphi(z) H_i(z)\, dz \qquad (i = 1, 2, 3, \ldots).$$

This solution is gotten by the introduction of $\sqrt{\varphi_0(z)}$ which serves to make all terms of the form $c_i\varphi_i(z) : \sqrt{\varphi_0(z)} = \sqrt{\varphi_0(z)}\ c_i H_i(z)$ ($i = 1, 2, 3 \ldots n$) orthogonal to each other in the interval $-\infty$ to $+\infty$.

In all the above expansions of a frequency series we have used the expression $\varphi_0(z) = e^{-z^2/2}$ as the generating function (see footnote on page 26), while as a matter of fact the true value of $\varphi_0(z)$ is given by the equation $\varphi_0(z) = e^{-z^2/2} : \sqrt{2\pi}$.

The definite integral on page 32

$$(-1)^i \int_{-\infty}^{+\infty} H_i(z)\varphi_i(z)\,dz = \lfloor i \int_{-\infty}^{+\infty} e^{-z^2:2} dz = \lfloor i \sqrt{2\pi}$$

will therefore have to be divided by $\sqrt{2\pi}$, and the value of the general coefficient c_i will henceforth be reduced to

$$c_i = \frac{\int_{-\infty}^{+\infty} \varphi(z)\, H_i(z)\,dz}{(-1)^i \lfloor i}$$

where $H_i(z)$ is the Hermite polynomial of order i defined by the relation

$$H_i(z) = z^i - \frac{i(i-1)}{2} z^{i-2} + \frac{i(i-1)(i-2)(i-3)}{2\cdot 4} z^{i-4} - \frac{i(i-1)(i-2)(i-3)(i-4)(i-5) z^{i-6}}{2\cdot 4\cdot 6} + \ldots$$

On this basis we obtain the following values for the first four coefficients :—

$$c_0 = \int_{-\infty}^{+\infty} \varphi(z)dz = 1$$

$$c_1 = (-1)^1 \int_{-\infty}^{+\infty} \varphi(z)z\,dz : \lfloor 1$$

$$c_2 = (-1)^2 \int_{-\infty}^{+\infty} (z^2 - 1)\,\varphi(z)\,dz : \lfloor 2$$

$$c_3 = (-1)^3 \int_{-\infty}^{+\infty} z^3 - 3z)\,\varphi(z)\,dz : \lfloor 3$$

$$c_4 = (-1)^4 \int_{-\infty}^{+\infty} (z^4 - 6z^2 + 3z)\,\varphi(z)\,dz : \lfloor 5$$

While the above development of an arbitrary frequency distribution has reference to $\varphi(z)$, or the relative frequency function, it is, however, equally well adapted to the representation of absolute frequencies as expressed by the function, $F(z)$. If N is the total number of individual observations, or in other words the area of the frequency curve, we evidently have

$$F(z) = N\varphi(z) \text{ or } \int_{-\infty}^{+\infty} F(z)dz = N\int_{-\infty}^{+\infty} \varphi(z)dz = N.$$

Since N is a constant quantity we may, therefore, write the expansion of $F(z)$ as follows:

$$F(z) = N\left[c_0\varphi_0(z) + c_1\varphi_1(z) + c_2\varphi_2(z) + \ldots\right] = \\ = N\sum c_i H_i(z) e^{-z^2:2}$$

where the coefficients c_i have the value

$$c_i = \frac{(-1)^i}{N\lfloor i} \int_{-\infty}^{+\infty} F(z) H_i(z)\, dz \text{ for } i = 1, 2, 3, \ldots$$

and where

$$N = \int_{-\infty}^{+\infty} F(z)\, dz.$$

Since all the Hermite functions are polynomials in z, it can be readily seen that the coefficients c may be expressed as functions of the power sums or of the previously mentioned symmetrical functions s, where

$$s_r = \int_{-\infty}^{+\infty} z^r F(z)\, dz.$$

These particular integrals originally introduced by Thiele in the development of the semi-invariants have been called by Pearson the "*moments*" of the frequency function, $F(z)$, and s_r is called the r^{th} moment of the variate z with respect to an arbitrary origin.

It can be readily seen that the moment of order zero, or s_0 is

$$s_0 = \int_{-\infty}^{+\infty} z^0 F(z)\,dz = N = N \int_{-\infty}^{+\infty} \varphi(z)\,dz.$$

Hence we have for the first coefficient c_0.

$$c_0 = \int_{-\infty}^{+\infty} F(z)\,dz : \int_{+\infty}^{+\infty} F(z)\,dz = 1.$$

We are, however, in a position to further simplify the expression for $F(z)$.

As already mentioned we are at liberty to choose arbitrarily both the origin and the unit of the Cartesian coordinate system for the frequency curve without changing the properties of this curve. Now by making a proper choice of the Cartesian system of reference we can make the coefficients c_1 and c_2 vanish. In order to obtain this object the origin of the system must be so chosen that

$$c_1 = \frac{-1}{\lfloor 1} \int_{-\infty}^{+\infty} z F(z)\,dz : \int_{-\infty}^{+\infty} F(z)\,dz = 0.$$

This means that the semi invariant $s_1 : s_0 = \lambda_1$ must vanish. It can be readily seen that the above expression for λ_1, is nothing more than the usual form for the mean value of a series of variates. Moreover, we know that the algebraic sum (or in the case of continuous variates, the integral) of the variates around the mean value is always

equal to zero. Hence by writing for z the expression $(z—M)$ when M equals the mean value or λ_1 we can always make c_1 vanish.

To attain our second object of making c_2 vanish we must choose the unit of the coordinate system in such a way that the expression

$$c_2 = \frac{(-1)^2}{\lfloor 2} \int_{-\infty}^{+\infty} F(z) H_2(z)\, dz : \int_{-\infty}^{+\infty} F(z)\, dz = 0$$

which implies that

$$\left[\int_{-\infty}^{+\infty} F(z) z^2\, dz - \int_{-\infty}^{+\infty} F(z)\, dz \right] : \int_{-\infty}^{+\infty} F(z)\, dz = 0$$

or that $s_2 : s_0 - 1 = 0$, or when expressed in terms of the semi-invariants that

$$\lambda_2 = (s_2 s_0 - s_1^2) : s_0^2 = 1.$$

But by choosing the mean as the origin of the system the term $s_1 : s_0$ is equal to 0 and we have therefore $\lambda_2 = \sigma^2 = s_2 : s_0 = 1$. Hence, by selecting as the unit of our coordinate system $\sqrt{\lambda_2}$ or σ, where σ is technically known as the dispersion or standard deviation of the series of variates, we can make the second coefficient c_2 vanish.

In respect to the coefficients c_3 and c_4 we have now

$$c_3 = \frac{(-1)^3}{\underline{|3}}\left[\int_{-\infty}^{+\infty} z^3 F(z)\,dz - 3\int_{-\infty}^{+\infty} zF(z)\,dz\right] : \int_{-\infty}^{+\infty} F(z)\,dz$$

which reduces to $-\frac{s_3 : s_0)}{\underline{|3}}$, while

$$c_4 = \frac{(-1)^4}{\underline{|4}}\left[\int_{-\infty}^{+\infty} z^4 F(z)\,dz - 6\int_{-\infty}^{+\infty} z^2 F(z)\,dz + 3\int_{-\infty}^{+\infty} F(z)\,dz\right] : \int_{-\infty}^{+\infty} F(z)\,dz$$

which reduces to

$$\left[\frac{s_4}{s_0} - \frac{6s_2}{s_0} + \frac{3s_0}{s_0}\right] : \underline{|4} = \left[\frac{s_4}{s_0} - 3\right] : \underline{|4}.$$

While the coefficients of higher order may be determined with equal ease, it will in general be found that the majority of moderately skew frequency distributions can be expressed by means of the first 4 parameters or coefficients.

11. COEFFICIENTS EXPRESSED AS SEMI-INVARIANTS We shall now show how the same results for the values of the coefficients may be obtained from the definition of the semi-invariants. Since we have proven that a frequency function, $F(z)$, may be expressed by the series

$$F(z) = \sum c_i \varphi_i(z)$$

we may from the definition of the semi-invariants write down the following identity:—

$$s_0 e^{\frac{\lambda_1 \omega}{\lfloor 1} + \frac{\lambda_2 \omega^2}{\lfloor 2} + \dots} = \\ = N \int_{-\infty}^{+\infty} e^{z\omega} (c_0 \varphi_0(z) + c_1 \varphi_1(z) + c_2 \varphi_2(z) + \dots) \, dz$$

where N is the area of the frequency curve.

The general term on the right hand side of the equation will be of the form

$$c_r \int_{-\infty}^{+\infty} e^{z\omega} \varphi_r(z) \, dz$$

where the integral may be evaluated by partial integration as follows:—

$$\int_{-\infty}^{+\infty} e^{z\omega} \varphi_r(z) \, dz = e^{z\omega} \varphi_{r-1}(z) \Big]_{-\infty}^{+\infty} - \omega \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-1}(z) \, dz,$$

and where the first term on the right vanishes leaving

$$\int_{-\infty}^{+\infty} e^{z\omega} \varphi_r(z) \, dz = (-\omega)^1 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-1}(z) \, dz.$$

Continuing in the same manner we obtain by successive integrations

$$(-\omega)^1 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-1}(z)\,dz = (-\omega)^2 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-2}(z)\,dz$$

$$(-\omega)^2 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-2}(z)\,dz = (-\omega)^3 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-3}(z)\,dz$$

. . . . . .

. . . . . .

from which we finally obtain the relation

$$\int_{-\infty}^{+\infty} e^{z\omega} \varphi_r(z)\,dz = (-\omega)^r \int_{-\infty}^{+\infty} e^{z\omega} \varphi_0(z)\,dz =$$

$$= \frac{(-\omega)^r}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{z\omega - \frac{z^2}{2}}\,dz.$$

This latter integral may be written as

$$\frac{(-\omega)^r}{\sqrt{2\pi}} e^{\frac{\omega^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z-\omega)^2}\,dz =$$

$$= \frac{(-\omega)^r}{\sqrt{2\pi}} e^{\frac{\omega^2}{2}} \sqrt{2\pi} = (-\omega)^r e^{\frac{\omega^2}{2}}.$$

Consequently the relation between the semi-invariants and the frequency function may be written as follows:—

$$s_0 e^{\frac{\lambda_1\omega}{\lfloor 1}+\frac{\lambda_2\omega^2}{\lfloor 2}+\cdots} = = N\left[c_0 - c_1\omega + c_2\omega^2 - c_3\omega^3 + \ldots\right] e^{\frac{\omega^2}{2}},$$

or

$$s_0 e^{\frac{\lambda_1\omega}{\lfloor 1}+\frac{\omega^2}{\lfloor 2}(\lambda_2-1)+\cdots} = = N\left[c_0 - c_1\omega + c_2\omega - c_3\omega^3 + \ldots\right].$$

By successive differentiation with respect to ω and by equating the coefficients of equal powers of ω we get in a manner similar to that shown on page 13 the following results :—

$$c_0 = \frac{s_0}{N} = \frac{s_0}{s_0} = 1$$

$$c_1 = -\lambda_1$$

$$c_2 = \frac{1}{\lfloor 2}\left[(\lambda_2 - 1) + \lambda_1^2\right]$$

$$c_3 = \frac{-1}{\lfloor 3}\left[\lambda_3 + 3(\lambda_2 - 1)\lambda_1 + \lambda_3^1\right]$$

$$c_4 = \frac{1}{\lfloor 4}\left[\lambda_4 + 4\lambda_3\lambda_1 + 3(\lambda_2 - 1)^2 + 6(\lambda_2 - 1)\lambda_1^2 + \lambda_1^4\right].$$

If we now again choose the origin at λ_1, or let $\lambda_1 = 0$, and choose $\sqrt{\lambda_2} = 1$ as the unit of our coordinate system we have :—

$$c_0 = 1,\ c_1 = 0,\ c_2 = 0,\ c_3 = \frac{-1}{\lfloor 3}\lambda_3,\ c_4 = \frac{1}{\lfloor 4}\lambda_4.$$

12. LINEAR TRANSFORMATION The theoretical development of the above formulae explicitly assumes that the variate, z, is measured in terms of the dispersion or $\sqrt{\lambda_2(z)}$ and with $\lambda_1(z)$ as the origin of the coordinate system. In practice the observations or statistical data are, however, invariably expressed with reference to an arbitrarily chosen origin (in the majority of cases the natural zero of the number scale) and expressed in terms of standard units, such as centimeters, grams, years, integral numbers, etc.

Let us denote the general variate in such arbitrarily selected systems of reference by x. Our problem then consists in transforming the various semi-invariants, $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ . . . . . to the z system of reference with $\lambda_1(z)$ as its origin and $\sqrt{\lambda_2(z)}$ as its unit. Such a transformation may always be brought about by means of the linear substitution

$$z = ax + b$$

which in a purely geometrical sense implies both a change of origin and unit. On page 16 we proved the following general properties of the semi-invariants

$$\lambda_1(z) = \lambda_1(ax+b) = a\lambda_1(x) + b$$
$$\lambda_r(z) = \lambda_r(ax+b) = a^r\lambda_r(x).$$

Let us now write $\lambda_1(x) = M$ and $\lambda_2(x) = \sigma^2$, we then have the following relations:—

$$\lambda_1(z) = aM + b$$
$$\lambda_2(z) = a^2\sigma^2.$$

Since the coordinate system of reference must be chosen in such a manner that $\lambda_1(z) = 0$ and $\sqrt{\lambda_2(z)} = 1$ we have:—

$$aM + b = 0$$
$$a\sigma = 1$$

from which we obtain $a = \frac{1}{\sigma}$ and $b = \frac{-M}{\sigma}$, which brings z on the form: $z = (x—M) : \sigma$ while $\varphi_0(z)$ becomes

$$\varphi_0(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-M)^2 : 2\sigma^2}.$$

Moreover, we have $\lambda_r(z) = \lambda_r(x) : \sigma^r$ for all values of $r \geq 2$. We are now able to epitomize the computations of the semi-invariants under the following simple rules.

(1) Compute $\lambda_1(x)$ in respect to an arbitrary origin. The numerical value of this parameter with opposite sign is the origin of the frequency curve.

(2) Compute $\lambda_r(x)$ for all values of $r \geq 2$. The numerical values of those parameters divided with $(\sqrt{\lambda_2(x)})^r$, or σ^r, for $r = 2, 3, 4, \ldots$ . . . . are the semi-invariants of the frequency curve.

13. CHARLIER'S SCHEME OF COMPUTATION

The general formulae for the semi-invariants were given on page 13. In practical work it is, however, of importance to proceed along systematic lines and to furnish an automatic check for the correctness of the computations. Several systems facilitating such work have been proposed by various writers, but the most simple and elegant is probably the one proposed by M. Charlier and which is shown in detail with the necessary control checks on the following page. Charlier employs moments, while we in the following demonstration shall prefer the use of the semi-invariants.

If we define the power sums of the relative frequencies $\varphi(x)$ by the relation

$$m_r = \int_{-\infty}^{+\infty} x^r F(x)\,dx : \int_{-\infty}^{+\infty} F(x)\,dx \quad (r = 0,\ 1,\ 2,\ 3, \ldots),$$

we find that the expressions for the semi-invariants as given on page 13 may be written as follows:—

$$\lambda_1 = m_1$$
$$\lambda_2 = m_2 - m_1^2$$
$$\lambda_3 = m_3 - 3m_2m_1 + 2m_1^3$$
$$\lambda_4 = m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$

The advantage of the Charlier scheme for the computation of the semi-invariants lies in the fact that it furnishes an automatic check of the final results. If we expand the expression $(x+1)^4 F(x)$ we have:—

$$x^4F(x)+4x^3F(x)+6x^2F(x)+4xF(x)+F(x)$$

or

$$\sum(x+1)^4F(x) = s_4+4s_3+6s_2+4s_1+s_0,$$

which serves as an independent control check of the computations. Moreover, another check is furnished by the relation

$$m_4 = \lambda_4+4m_1\lambda_3+6m_1^2\lambda_2+3\lambda_2^2+m_1^4.$$

In order to illustrate the scheme we choose the following age distribution of 1130 pensioned functionaries in a large American Public Utility corporation.

Ages	No. of Pensioners	Ages	No. of Pensioners
35—39	1	65—69	286
40—44	6	70—74	248
45—49	17	75—79	128
50—54	48	80—84	38
55—59	118	85—89	13
60—64	224	over 90	3

The complete calculations of the coefficients c are shown in the appended scheme by Charlier.

x	$F(x)$	$xF(x)$	$x^2F(x)$	$x^3F(x)$	$x^4F(x)$	$(x+1)^4F(x)$
35—39..... —6	1	6	36	216	1296	625
40—44..... —5	6	30	150	750	3750	1536
45—49..... —4	17	68	272	1088	4352	1377
50—54..... —3	48	144	432	1296	3888	768
55—59..... —2	118	236	472	944	1888	118
60—64..... —1	224	224	224	224	244	0
65—69..... 0	286	0	0	0	0	286
	700	708	1586	4522	15398	4710
70—74..... +1	248	248	248	248	248	3968
75—79..... +2	128	256	512	1024	2048	10368
80—84..... +3	38	114	342	1026	3078	9728
85—89..... +4	13	52	208	832	3328	8125
90—94..... +5	2	10	50	250	1250	2592
95—99..... +6	1	6	36	216	1296	2401
	430	686	1396	3596	11248	37182
s_r =	1130	—22	2982	—922	26646	41892
m_r =	1.0000	—.0195	2.6378	—.8156	23.5699	

$\lambda_1 = m_1 = -.0195$
$\lambda_1^2 = m_1^2 = .0004$
$\lambda_1^3 = m_1^3 = .0000$
$\lambda_1^4 = m_1^4 = .0000$

$m_2 = 2.6378$
$-m_1^2 = -.0004$
$\lambda_2 = 2.6374 = \sigma^2$

$\sqrt{\lambda_2} = 1.6240 = \sigma$
$\sigma^3 = 4.2831$
$\sigma^4 = 6.9558$

Control Check
$s_4 = 26646$
$4s_3 = -3688$
$6s_2 = 17892$
$4s_1 = -88$
$s_0 = 1130$
41892

$m_2 m_1 = -.0513,\ m_3 m_1 = .0159,\ m_2^2 = 6.9580,\ m_2 m_1^2 = .0010$

$m_3 = -.8156$
$-3m_2 m_1 = .1539$
$2m_1^3 = .0000$
$\lambda_3 = -.6617$

$c_3 = \lambda_3 : \sigma^3 = -.1545$
$-c_3 : \underline{|3} = .0258$

$m_4 = 23.5699$
$-m_3 m_1 = -.0636$
$-3m_2^2 = -20.8740$
$12 m_2 m_1^2 = .0127$
$6m_1^4 = .0000$
$\lambda_4 = 2.6450$

$c_4 = \lambda_4 : \sigma^4 = .3803$
$c_4 : \underline{|4} = .0158$

Control Check
$\lambda_4 = 2.6450$
$4m_1\lambda_3 = .0516$
$6m_1^2\lambda_2 = .0060$
$3\lambda_2^2 = 20.8677$
$m_1^4 = .0000$
$23.5703 = m_4$

The above computations give the numerical values of the frequency function which now may may be written as follows:

$$F(x) = 1130\,[(\varphi_0(x) + .0258\,\varphi_3(x) .0158\,\varphi_4(x)]$$

where

$$\varphi_0(x) = \frac{1}{1.624\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x+.0195}{1.6240}\right)^2}$$

14. COMPARISON BETWEEN OBSERVED DATA AND THEORETICAL VALUES

The next step is now to work out the numerical values of $F(x)$ for various values of x and compare such values with the ones originally observed. This process is shown in detail in the following scheme.

Column (1) gives the values of the variate x reckoned from the provisional origin, or the centre of the age interval 65-69. (2) is x less the first semi-invariant, whereby the origin is shifted to the mean or λ. Column (3) represents the final linear transformation: $z = (x - \lambda_1) : \sigma$.

Columns (4), (5) and (6) are copied directly from the standard tables of Jørgensen or Charlier. Column (7) is (5) multiplied by 0.0258 or the product $-[c_3\varphi_3(z)] : \underline{|3}$, while (8) is $[c_4\varphi_4(z)] : \underline{|4}$.

Column (9) is the sum of (4), (7) and (8). If we now distribute the area $N = s_0$ or 1130 *pro*

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Obs.
x	$x-\lambda_1$	$(x-\lambda_1):\sigma$	$\varphi_0(z)$	$\varphi_3(z)$	$\varphi_4(z)$					
−7	−6.98	−4.300	.0001	+.0058	+.0170	+.0001	+.0003	.0005		
−6	5.98	3.682	.0005	.0176	.0479	.0005	.0008	.0018	1	1
−5	4.98	3.067	.0036	.0710	.1267	.0018	.0020	.0074	5	6
−4	3.98	2.451	.0198	.1458	+.0602	.0038	+.0009	.0245	17	17
−3	2.98	1.835	.0741	+.0500	−.4345	+.0013	−.0068	.0686	48	48
−2	1.98	1.219	.1987	−.3502	−.7036	−.0090	.0111	.1696	118	118
−1	−0.98	−0.603	.3326	−.5287	+.3160	−.0136	+.0050	.3240	226	224
0	+0.02	+0.012	.3989	+.0143	1.1963	+.0004	+.0189	.4182	291	286
+1	1.02	0.628	.3273	.5359	+.2584	.0138	+.0041	.3452	241	248
+2	2.02	1.244	.1895	+.3325	−.7157	+.0086	−.0113	.1808	126	128
+3	3.02	1.860	.0707	−.0605	−.4094	−.0015	−.0065	.0627	44	38
+4	4.02	2.475	.0186	.1443	+.0703	.0037	+.0011	.0212	15	13
+5	5.02	3.091	.0034	.0680	.1241	.0018	.0020	.0036	3	2
+6	6.02	3.707	.0004	.0165	.0456	.0004	.0007	.0007	1	1
+7	+7.02	+4.322	.0001	−.0050	.0162	−.0001	+.0003	.0003		

rata according to (9), we finally reach the theoretical frequency distribution expressed in 5-year age intervals and shown in column (10) alongside which we have inserted the originally observed values. Evidently the fit is satisfactory. It will be noted that the final frequency series is expressed in units of 5-year age intervals. This, however, is only a formal representation. By subdividing the unit intervals of column (1) in 5 equal parts, and by computing all the other columns accordingly, we get the theoretical frequency series expressed in single year age intervals.

15. THE PRINCIPLE OF METHOD OF LEAST SQUARES The following paragraph purports to give a brief exposition of the determination of the coefficients in the Gram or Laplacean—Charlier series in the sense of the method of least squares as a strict problem of maxima and minima, wholly independent of the connection between the method of least squares and the error laws of precision measurements. [1]

The simple problem in maxima and minima which forms the fundamental basis of the method

[1] In the following demonstration I am adhering to the brief and lucid exposition of the Argentinean actuary, U. Broggi, in his exellent *Traite d' Assurances sur la Vie.*

of least squares is the following: Let m unknown quantities be determined by observations in such a manner that they are not observed directly but enter into certain *known* functional relations, $f_i(x_1, x_2, x_3, \ldots x_m)$, containing the unknown independent variables, $x_1, x_2, x_3, \ldots x_m$. Let furthermore the number of observations on such functional relations be n (where n is greater than m). The problem is then to determine the most plausible system of the values of the unknowns from the observed system.

$$\begin{aligned}
f_1(x_1, x_2, x_3, \ldots x_m) &= o_1 \\
f_2(x_1, x_2, x_3, \ldots x_m) &= o_2 \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
f_n(x_1, x_2, x_3, \ldots x_m) &= o_n
\end{aligned}$$

when $f_1, f_2, \ldots f_n$ are the known functional relations and $o_1, o_2, \ldots o_n$ their observed values. Such equations are known as *observation equations*.

In order to further simplify our problem we shall also assume that

1 All the equations of the system have the same weight, and

2 All the equations are reduced to linear form.

By these assumptions the problem is reduced to find m unknowns from n linear equations.

$$
\begin{aligned}
a_1 x_1 + b_1 x_2 + \ldots &= o_1 \\
a_2 x_1 + b_2 x_2 + \ldots &= o_2 \\
a_3 x_1 + b_3 x_2 + \ldots &= o_3 \\
\ldots\ldots\ldots\ldots\ldots & \\
\ldots\ldots\ldots\ldots\ldots & \\
a_n x_1 + b_n x_2 + \ldots &= o_n
\end{aligned}
$$

Since n is greater than m we find the problem over-determined, and we therefore seek to determine the unknown quantites, x_1, x_2, . . . x_m in such a way that the sum of the squares of the differences between the functional relations and the observed values, o becomes a minimum. This implies that the expression

$$\sum_{i=1}^{i=m} (a_i x_1 + b_i x_2 + \ldots - o_i)^2 = \psi(x_1, x_2, \ldots x_m)$$

must be a minimum or the simultaneous existence of the equations.

$$\frac{\delta \psi}{\delta x_1} = 0, \quad \frac{\delta \psi}{\delta x_2} = 0, \ldots \frac{\delta \psi}{\delta x_m} = 0. \qquad (I)$$

If we now introduce the following notation

$$a_i x_1 + b_i x_2 + \ldots - o_i = \lambda_i \text{ for } i = 1, 2, 3, \ldots n,$$

the m equations in the above system *(I)* evidently take on the following form

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n = 0$$
$$\lambda_1 b_1 + \lambda_2 b_2 + \ldots + \lambda_n b_n = 0$$
$$\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots$$

If we now again re-substitute the expressions for λ in terms of the linear relations

$$a_i x_1 + b_i x_2 + \ldots o_i = \lambda_i, \text{ for } i = 1, 2, 3, \ldots n,$$

and collect the coefficients of x_1, x_2, $\ldots x_n$, these equations may be expressed in the following symbolical form:

$$[aa]x_1 + [ab]x_2 + \ldots\ldots - [ao] = 0$$
$$[ab]x_1 + [bb]x_2 + \ldots\ldots - [bo] = 0$$
$$\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots$$
$$[ak]x_1 + [bk]x_2 + \ldots\ldots + [kk]x_m - [ko] = 0$$

where $[aa] = a_1^2 + a_2^2 + \ldots\ldots$
$[ab] = a_1 b_1 + a_2 b_2 + \ldots\ldots$

is the Gaussian notation for the homogeneous sum products.

The above equations are known as *normal equations*, and it is readily seen that there is one normal equation corresponding to each unknown. Our problem is therefore reduced to the solution of a system of simultaneous linear equations of m

unknowns. If m is a small number, or, what amounts to the same thing, there are only two or three unknowns the solution can be carried on by simple algebraic methods or determinants. If the number of unknowns is large these methods become very laborious and impractical. It is one of the achievements of the great German mathematician, Gauss, to have given us a method of solution which reduces this labor to a minimum and which proceeds along well defined systematic and practical lines. The method is known as the Gaussian algorithmus of successive elimination.

16. GAUSS' SOLUTION OF NORMAL EQUATIONS

For the sake of simplicity we shall limit ourselves to a system of four normal equations of the form

$$\begin{aligned}
[aa]x_1 + [ab]x_2 + [ac]x_3 + [ad]x_4 - [ao] &= 0 \\
[ab]x_1 + [bb]x_2 + [bc]x_3 + [bd]x_4 - [bo] &= 0 \\
[ac]x_1 + [bc]x_2 + [cc]x_3 + [cd]x_4 - [co] &= 0 \\
[ad]x_1 + [bd]x_2 + [cd]x_3 + [dd]x_4 - [do] &= 0
\end{aligned}$$

The generalization to an arbitrary number of unknowns offers no difficulties, however.

On account of their symmetrical form the above equations may also be written in the more convenient form, viz.:

$$\begin{aligned}
[aa]x_1 + [ab]x_2 + [ac]x_3 + [ad]x_4 - [ao] &= 0 \\
[bb]x_2 + [bc]x_3 + [bd]x_4 - [bo] &= 0 \\
[cc]x_3 + [cd]x_4 - [co] &= 0 \\
[dd]x_4 - [do] &= 0
\end{aligned}$$

From the first equation we find

$$x_1 = \frac{[ao]}{[aa]} - \frac{[ab]}{[aa]}x_2 - \frac{[ac]}{[aa]}x_3 - \frac{[ad]}{[aa]}x_4.$$

Substituting this value in the following equations and by the introduction of the new symbol

$$[ik] - \frac{[ai]}{[aa]}[ak] = [ik.1]$$

we now obtain a new system of equations of a lower order and of the form

$$\begin{aligned}
[bb.1]x_2 + [bc.1]x_3 + [bd.1]x_4 - [bo.1] &= 0 \\
[cc.1]x_3 + [cd.1]x_4 - [co.1] &= 0 \\
[dd.1]x_4 - [do.1] &= 0
\end{aligned}$$

Solving for x_2 we have

$$x_2 = \frac{[bo.1]}{[bb.1]} - \frac{[bc.1]}{[bb.1]}x_3 - \frac{[bd.1]}{[bb.1]}x_4.$$

Substituting in the following equations and writing

$$[ik.1] - \frac{[bc.1]}{[bb.1]}[bk.1] = [ik.2]$$

we have

$$[cc.2]\,x_3 + [cd.2]\,x_4 = [co.2]$$
$$[dd.2]\,x_4 = [do.2]$$

or

$$x_3 = \frac{[co.2]}{[cc.2]} - \frac{[cd.2]}{[cc.2]}\,x_4.$$

Moreover, by writing

$$[ik.2] = [ci.2]\,\frac{[ck.2]}{[cc.2]} = [ik.3],$$

we have finally

$$[dd.3]\,x_4 = [do.3]$$

This gives us the final reduced normal equation of the lowest order. By successive substitution we therefore have:

$$x_4 = \frac{[do.3]}{[dd.3]}$$

$$x_3 = \frac{[co.2]}{[cc.2]} - \frac{[cd.2]}{[cc.2]}\,x_4$$

$$x_2 = \frac{[bo.1]}{[bb.1]} - \frac{[bc.1]}{[bb.1]} - \frac{[bd.1]}{[bb.1]}$$

$$x_1 = \frac{[ao]}{[aa]} - \frac{[ab]}{[aa]}\,x_2 - \frac{[ac]}{[aa]}\,x_3 - \frac{[ad]}{[aa]}\,x_4$$

as the ultimate solution of the unknowns.

17. ARITHMETICAL APPLICATION OF METHOD The example in paragraph 13 gave an illustration of the application of the method of moments. As previously stated this method works quite well in cases of moderate skewness, but is less successful in extremely skew curves and where the excess is large. We shall now give an illustration of the calculation of the parameters by the method of least squares. The example we choose is the well-known statistical series by the distinguished Dutch botanist, de Vries, on the number of petal flowers in *Ranunculus Bulbosus*. This is also one of the classical examples of Karl Pearson in his celebrated original memoirs on skew variation. Although the observations of de Vries lend themselves more readily to the method of logarithmic transformation, which we shall discuss in a following chapter, we have deliberately chosen to use it here for two specific reasons. Firstly it is a most striking illustration in refutation of the immature criticism of the Gram-Charlier series by a certain young and very incautious American actuary, Mr. M. Davis, who has gone on record with the positive statement, "that the Charlier series fails completely in case of appreciable skewness". Secondly (and this is the more important reason) it offers an excellent drill for the student in the practical applications of the method of least

squares because it gives in a very brief compass all the essential arithmetical details. The observations of de Vries are as follows:

No. of petals	x	$F(x) = o_x$
5	0	133
6	1	55
7	2	23
8	3	7
9	4	2
10	5	2

where $F(x)$ denotes the absolute frequencies. The observed frequency distribution is well nigh as skew as it can be and represents in fact a one-sided curve, and should therefore—if the statement by Mr. Davis is correct—show an absolute defiance to a graduation by the Gram-Charlier series.

The process we shall use in the attempted mathematical representation of the above series is a combination of the method of semi-invariants and the method of least squares. Following Thiele's advice we determine the first two semi-invariants in the generating function directly from the observations while the coefficients of this function and its derivations are determined by the least square method.

Choosing the provisional origin at 5, we obtain the following values for the crude moments.

$$s_0 = 222,\ s_1 = 140,\ s_2 = 292,\ s_3 = 806,\ s_4 = 2{,}752,$$
$$s_5 = 10{,}790,\ s_6 = 46{,}072,\ s_7 = 207{,}226,$$

from which we find that

$$\lambda_0 = 1,\ \lambda_1 = 0.631,\ \lambda_2 = 0.917,\ \lambda_3 = 1.644,$$
$$\lambda_4 = 3.377,\ \lambda_5 = 5.972,\ \lambda_6 = -2.911,$$
$$\lambda_7 = 122.638.$$

All these semi-invariants with the exception of the two first are, however, so greatly influenced by random sampling in the small observation series that it is hopeless to use them in the determination of the constants in the Gram-Charlier series. In fact an actual calculation does not give a very good result beyond that of a first rough approximation. The generating function, on the other hand, may be expressed by the aid of the two first semi-invariants as follows:

$$\varphi_0(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2:2},$$

where z is given by the linear transformation:

$$z = (x - 0.631) : 0.9576. \qquad (\sqrt{\lambda_2} = 0.9576).$$

We now propose to express the observed function $F(x)$ or $\varphi(z)$ by a Gram-Charlier series of the form:

$$F(x) = \varphi(z) = k_0\varphi_0(z) + k_3\varphi_3(z) + k_4\varphi_4(z).$$

In this equation we know the values of the generating function and its derivatives for various values of the variate z as found in the tables of Jørgensen and Charlier, while the quantities k are unknowns. On the other hand we know 6 specific values of $F(x)$ as directly observed in de Vries's observation series. We are thus dealing with a system of typical linear observation equations of the forms described in paragraphs 15 and 16 and which lend themselves so admirably to the treatment by the method of least squares.

From the above linear relation between x and z we can directly compute the following table for the transformed variate z.

x	z
0	—0.688
1	+0.402
2	+1.493
3	+2.583
4	+3.674
5	+4.764.

The numerical values of $\varphi_0(z)$ and its derivatives as corresponding to the above values of z can be taken directly from the standard tables of Jørgensen and Charlier. We may therefore write down the following observation equations:

φ_0	φ_3	φ_4	o
$.3148k_0$	$-.5472k_3$	$+.1207k_4$	$-133 = 0$
$.3679k_0$	$+.4198k_3$	$+.7566k_4$	$- 55 = 0$
$.1308k_0$	$+.1506k_3$	$-.7073k_4$	$- 23 = 0$
$.0145k_0$	$-.1346k_3$	$+.1062k_4$	$- 7 = 0$
$.0005k_0$	$-.0180k_3$	$+.0486k_4$	$- 2 = 0$
$.0001k_0$	$-.0005k_3$	$+.0020k_4$	$- 2 = 0$

for which we now propose to determine the unknown values of k by the least square method.

While this method may of course be applied directly to the above data, it will generally be found of advantage to start with some approximate values of the k's. It is found in practice that this approximate step saves considerable labour in the formation and ultimate solution of the normal equations.

Although the first approximation in the case of numerous unknowns must be in the nature of a more or less shrewd guess, which facility can only be attained by constant practice in routine mathematical computing, we are, however, in this specific instance able to tell something about the nature o fthe coefficients from purely *a priori* considerations. We know for instance from the form of the Gram-Charlier series that the coefficient k_0 of the generating function must be nearly equal to the area of the curve, which in this particular instance is 222. Moreover, a mere glance at the observed series tells us that it has a decidedly

large skewness in negative direction from the mean coupled with a tendency of being "top heavy", indicating positive excess. We can therefore assume as a first approximation that the coefficients of the derivatives of uneven order are negative and the coefficients of derivatives of even order are positive.

From such purely common sense *a priori* considerations we therefore guess the following first approximations, viz.:

$$k_0^1 = 222,\ k_3^1 = -25,\ k_4^1 = 30.$$

The probable values of the various k's may be written as

$$k_i = r_i k_i^1 \text{ for } i = 0, 3, 4,$$

and our problem is therefore to find the correction factor r with which the approximate value k_i^1 must be multiplied so as to give k_i.

Applying the various values of k_i^1 to the original observation equations on page 64 we obtain the following schedule for the numerical factors of r_i.

a	*b*	*c*	*o*	*s*
69.9	+13.7	+ 3.6	—133.0	—45.8
81.7	—10.5	22.7	— 55.0	+38.9
29.1	— 3.8	—21.2	— 23.0	—18.9
3.3	+ 3.4	+ 3.2	— 7.0	+ 2.9
0.1	+ 0.5	+ 1.5	— 2.0	+ 0.1
0.0	+ 0.0	+ 0.0	— 2.0	— 2.0
184.1	+ 3.3	+ 9.8	—222.0	—24.8

where the additional control column s serves as a check.

The subsequent formation of the various sum-products and normal equations is shown in the following schedules together with the s columns as a check.

aa	ab	ac	ao	as
+ 4,886	+958	+ 252	— 9,297	—3201
+ 6,675	—858	+1,855	— 4,494	+3178
+ 847	—111	— 617	— 669	— 550
+ 11	+ 11	+ 11	— 23	+ 10
+ 0	+ 0	+ 0	— 0	+ 0
+ 0	+ 0	+ 0	— 0	+ 0
+12,419	+ 0	+1,501	—14,483	— 563

bb	bc	bo	bs
+188	+ 49	— 1,822	— 628
+110	— 238	— 578	— 408
+ 14	+ 81	+ 87	+ 72
+ 12	+ 11	— 24	+ 10
+ 0	+ 0	— 1	+ 0
+ 0	+ 0	— 0	+ 0
+324	— 96	— 1,182	— 954

cc	co	cs
+ 13	— 479	— 165
+ 515	— 1,249	+ 883
+ 449	+ 488	+ 401
+ 10	— 22	+ 9
+ 2	— 3	+ 1
+ 0	+ 0	+ 0
+ 989	— 1,265	+1129

We may now write the normal equations in schedule form as follows:

ORIGINAL NORMAL EQUATIONS

(*a*)	+ 12,419	+	0	+	1501	—	14483
(1)		+	**0**	+	**0**	—	**0**
(*b*)		+	324	—	96	—	1182
(2)				+	**181**	—	**1750**
(*c*)				+	989	—	1265
(3)			**+.00000**		**+.12086**		**—1.16617**

The sum-products from the observation equations are shown in the rows marked (*a*), (*b*), (*c*). The row marked (3) and printed in italics is formed by dividing each of the figures in row (*a*) with 12,419. The row marked (1) contains the products of the figures in row (*a*) multiplied with the factor .00000. All these products happen in this case to be equal to zero. Row (2) is the products of the factor 0.12086 and the figures in row (*a*).

We next subtract row (1) from row (*b*), row (2) from row (*c*), which results in the following schedule, which is known as the first *reduction equation.*

FIRST REDUCTION EQUATIONS

(*a*)	+324	—	96	—	1182
(1)		+	**28**	+	**350**
(*b*)		+	808	+	485
(2)			**—.29626**		**—3.64814**

The above equations are treated in a similar manner as the original normal equations, and we have therefore the 2nd reduction equation of the form:

SECOND REDUCTION EQUATION

$+780 \qquad +135$

The solution for the unknown r's may now be shown as follows:

$r_4 = -135 : 780 = -.17308$

$r_3 = 3.64814 - (-.29626)\ (-.17308) = 3.59637$

$r_0 = 1.16617 - (0.0)\ 3.59637) - (.12086)$
$(-.17308) = 1.18709.$

From which we find:—

$$k_0 = 263.5, \qquad k_3 = -89.9, \qquad k_4 = -5.1$$

Applying these factors to the values of $\varphi_0(z)$, $\varphi_3(z)$ and $\varphi_4(z)$ we obtain the following result:—[1]

$k_0\varphi_0$	$k_3\varphi_3$	$k_4\varphi_4$	$\Sigma k_i\varphi_i$	Obs.
82.9	+49.2	—0.6	131.5	133
96.9	—37.7	—3.9	55.3	55
34.5	—13.5	+3.6	24.6	23
3.8	+12.1	—0.5	15.4	7
0.1	+ 1.0	—0.2	0.9	2
0.0	+ 0.0	—0.0	0.0	2

[1] For a closer approximation see my *Mathematical Theory of Probabilities* (Second Edition, New York, 1921).

18. TRANSFORMATION OF THE VARIATE

While it is always possible to express all frequency curves by an expansion in Hermite polynomials, the numerical labor when carried on by the method of least squares often involves a large amount of arithmetical work if we wish to retain more than four or five terms of the series. Other methods lessening the arithmetical work and making the actual calculations comparatively simple have been offered by several authors and notably by Thiele, who in his works discusses several such methods. Among those we may mention the method of the so-called free functions and orthogonal substitution, the method of correlates and the adjustment by elements. The chapters on these methods in Thiele's work are among some of the most important, but also some of the most difficult in the whole theory of observations and have not always been understood and appreciated by the mathematicians, chiefly on account of Thiele's peculiar style of writing. A close study of the Danish scholar's investigations is, however, well worth while, and Thiele's work along these lines may still in the future become as epochmaking in the theory of probability as some of the researches of the great Laplace. The theory of infinite determinants as used by M. Fredholm in the solution of integral equations is

another powerful tool which offers great advantages in the way of rapid calculation. All these methods require, however, that the student must be thoroughly familiar with the difficult theory upon which such methods rest, and they have for this reason been omitted in an elementary work such as the present treatise.

We wish, however, to mention another method which in the majority of cases will make it possible to employ the Gram or Laplacean—Charlier curves in cases with extreme skewness or excess. We have here reference to the method of logarithmic transformation of the variate, x.

19. THE GENERAL THEORY OF TRANSFORMATION One of the simplest transformations is the previously mentioned linear transformation of the form $z = f(x) = ax + b$, by which we can make two constants, c_1 and c_2 vanish. Other transformations suggest themselves, however, such as $f(x) = ax^2 + bx + c$, $f(x) = \sqrt{x}$, $f(x) = \log x$ and so forth. For this reason I propose to give a brief development of the general method of transformations of the statistical variates, mainly following the methods of Charlier and Jørgensen.

Stated in its most general form our problem

is: If a frequency curve of a certain variate is given by $F(x)$ what will be the frequency curve of a certain function of x, say $f(x)$?

The equation of the frequency curve is $y = F(x)$, which means that $F(x)\,dx$ is the probability that x falls in the interval between $x - \frac{1}{2}dx$ and $x + \frac{1}{2}dx$. The probability that a new variate z after the transformation $z = f(x)$, or $\chi(z) = x$, falls in the interval $z - \frac{1}{2}dz$ and $z + \frac{1}{2}dz$ is therefore simply

$$F[\chi(z)]\chi^1(z)\,dz = F(x)\,dx,$$

which gives in symbolic form the equation of the transformed frequency curve.

The frequency for $z = \mathrm{f}(x)$ is of course the same as for x. The ordinates of the frequency curve, or rather the areas between corresponding ordinates, are therefore not changed, but the abcissa axis is replaced by $f(x)$. Equidistant intervals of x will therefore not as a rule—except in the linear transformation—correspond to equidistant intervals of $f(x)$.

If, for instance, the frequency curve $F(x)$ is the Laplacean normal curve

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2 : 2\sigma^2}$$

and if we let $z = f(x) = x^2$ or $x = \sqrt{z}$, we have evidently

$$F(z) = \frac{1}{\sigma\sqrt{2\pi}} \frac{e^{-z:2\sigma^2}}{2\sqrt{z}}.$$

20. LOGARITHMIC TRANSFORMATION Of the various transformations the logarithmic is of special importance. It happens that even if the variate x forms an extremely skew frequency distribution its logarithms will be nearly normally distributed.

This fact was already noted by the eminent German psychologist, Fechner, and also mentioned by Bruhns in his *Kollektivmasslehre*. But neither Fechner nor Bruhns have given a satisfactory theoretical explanation of the transformation and have limited themselves to use it as a practical rule of thumb.

Thiele discusses the method under his adjustment by elements, but in a rather brief manner. The first satisfactory theory of logarithmic transformation seems to have been given first by Jørgensen and later on by Wicksell.[1]) Jørgensen

[1] The law of errors, leading to the geometric mean as the most probable value of the variate as discovered by Prof. Dr. Th. N. Thiele in 1867 may, however, be considered as a forerunner of Jørgensen's work.

first begins with the transformation of the normal Laplacean frequency curve. Letting $z = logx$ and bearing in mind that the frequency of x equals that of $logx$ we have

$$z = f(x) = \log x, \text{ or } x = \chi(z) = e^z \text{ and } dx = e^z dz.$$

The continuous power sums or moments of the *rth* order around the lower limit take on the form

$$(n\sqrt{2\pi})^{-1} N \int_0^{\infty} x^r e^{\frac{1}{2}\left(\frac{\log x - m}{n}\right)^2} dx =$$

$$= (n\sqrt{2\pi})^{-1} N \int_{-\infty}^{+\infty} e^{rz} e^{\frac{1}{2}\left(\frac{z-m}{n}\right)^2} e^z dz.$$

on the assumption the $logx$ is normally distributed.

The change in the lower limit in the second integral from $-\infty$ to zero arises simply from the fact that the logarithm of zero equals minus infinity and the point $-\infty$ is thus by the transformation moved up to zero.

By a straightforward transformation we may write the above integral as

$$M_r = \frac{N}{\sqrt{2\pi}} e^{m(r+1)+\frac{1}{2}n^2(r+1)^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt =$$

$$= N e^{m(r+1)+\frac{1}{2}n^2(r+1)^2}.$$

Changing from moments to semi-variants by means of the well-known relations

$$\begin{aligned}
\lambda_0 &= M_0 \\
\lambda_1 &= M_1 : M_0 \\
\lambda_2 &= (M_2 M_0 - M_1^2) : M_0^2 \\
\lambda_3 &= (M_3 M_0^2 - 3 M_2 M_1 M_0 + 2 M_1^3) : M_0^3 \\
\lambda_4 &= (M_4 M_0^3 - 4 M_3 M_1 M_0^2 - 3 M_2^2 M_0^2 + \\
&\quad + 12 M_2 M_1^2 M_0 - 6 M_1^4) : M_0^4 \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

we have

$$\begin{aligned}
\lambda_0 &= N e^{m + \frac{1}{2}n^2} \\
\lambda_1 &= e^{m + 1.5n^2} \\
\lambda_2 &= e^{2m + 3n^2} (e^{n^2} - 1) \\
\lambda_3 &= e^{3m + 4.5n^2} (e^{3n^2} - 3e^{n^2} + 2) \\
\lambda_4 &= e^{4m + 6n^2} (e^{6n^2} - 4e^{3n^2} - 3e^{2n^2} + 12e^{n^2} - 6). \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

These equations give the semi-invariants expressed in terms of m and n. On the other hand if we know the semi-invariants from statistical data or are able to determine these semi-invariants by *a priori* reasoning we may find the parameters m and n.

21. THE MATHEMATICAL ZERO

A point which we must bear in mind is that the above semi-invariants on account of the transformation are calculated around a zero point which corresponds to a fixed lower limit of the observations.

Very often the observations themselves indicate such a lower limit beyond which the frequencies of the variate vanish. In the case of persons engaged in factory work there is in most countries a well-defined legal age limit below which it is illegal to employ persons for work. Another example is offered in the number of alpha particles radiated from certain radioactive metals. Since the number of particles radiated in a certain interval of time must either be zero or a whole positive number it is evident that—1 must be the lower limit because we can have no negative radiations. Analogous limits exist in the age limit for divorces and in the amount of moneys assessed in the way of income tax.

The lower limit allows, however, of a more exact mathematical determination by means of the following simple considerations. It is evident that this lower limit must fall below the mean value of the frequency curve. Let us suppose that it is located at a point, a, located say η units in negative direction from the mean, $M = \lambda_1$, and let us to begin with select λ_1 as the origin of the coordinate system in which case the first semi-invariant, λ_1, is equal to zero. Transferring the origin to a the first semi-invariant equals η, while the semi-invariants of higher order remain the same as before the transformation and we have:

$$\lambda_1 - a = \eta = e^{m+1.5n^2}$$

$$\lambda_2 = \eta^2(e^{n^2} - 1) \text{ or } e^{n^2} = 1 + \lambda_2 : \eta^2$$

$$\lambda_3 = \eta^3\left[\left(\frac{\lambda_2}{\eta^2} + 1\right)^3 - 3\left(\frac{\lambda_2}{\eta^2} + 1\right) + 2\right] = \eta^3\left[\frac{\lambda_2^3}{\eta^6} + \frac{3\lambda_2^2}{\eta^4}\right]$$

which reduces to $\lambda_3\eta^3 - 3\lambda_2^2\eta^2 - \lambda_2^3 = 0$.

The solution of this cubic equation which has one real and two imaginary roots gives us the value of η or $\lambda_1 - a$ and thus determines the mathematical zero or lower limit. We have in fact:

$$n^2 = \log(1 + \lambda_2 : \eta^2) \text{ and}$$

$$m = \log \eta - 1.5n^2, \text{ while}$$

$$N = \lambda_0 : e^{m+1/2n^2}$$

22. LOGARITHMICALLY TRANSFORMED FREQUENCY SERIES

We have already shown that the generalized frequency curve could be written as

$$F(x) = c_0\varphi_0(x) - \frac{c_1\varphi_1(x)}{1!} + \frac{c_2\varphi_2(x)}{2!} - \frac{c_3\varphi_3(x)}{3!} + \ldots$$ [1]

where the Laplacean probability function

$$\varphi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-M)^2}{2\sigma^2}}$$

is the generating function with M and σ as its parameters.

The suggestion now immediately arises to use an analogous series in the case of the logarithmic transformation. In this case the frequency curve, $F(x)$, with a lower limit would be expressed as follows:

$$F(x) = k_0\Phi_0(x) - \frac{k_1\Phi_1(x)}{1!} + \frac{k_2\Phi_2(x)}{2!} - \frac{k_3\Phi_3(x)}{3!} + \ldots$$

while the generating function now is

$$\Phi_0(x) = \frac{1}{n\sqrt{2\pi}} e^{\frac{1}{2}\left[\frac{\log x - m}{n}\right]^2}$$

where m and n are the parameters.

---

[1] $n! = \lfloor n$.

Using the usual definition of semi-invariants we then have

$$s_0 e^{\frac{\lambda_1\omega}{1!}+\frac{\lambda\omega_2}{2!}+\frac{\lambda_3\omega_3}{3!}+\cdots} = s_0 + \frac{s_1\omega}{1!} + \frac{s_2\omega^2}{2!} + \frac{s_3\omega^3}{3!} + \cdots$$

$$= \int_0^\infty e^{x\omega}\left[k_0\Phi_0(x) - \frac{k_1\Phi_1(x)}{1!} + \frac{k_2\Phi_2(x)}{2!} - \right.$$

$$\left. - \frac{k_3\Phi_3(x)}{3!} + \cdots\right] dx.$$

The general term on the right hand side integral is of the form

$$(-1)^s k_s : s! \int_0^\infty e^{x\omega}\Phi_s(x)\,dx$$

where the integral may be evaluted by partial integration as follows:

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx = e^{x\omega}\Phi_{s-1}(x)\Big]_0^\infty - \omega\int_0^\infty e^{x\omega}\Phi_{s-1}(x)\,dx.$$

Since both $\Phi(x)$ and all its derivatives are supposed to vanish for $x = 0$ and $x = \infty$ the first term to the right becomes zero and

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx = -\omega\int_0^\infty e^{x\omega}\Phi_{s-1}(x)\,dx.$$

By successive integrations we then obtain the following recursion formula

$$(-\omega)^1 \int_0^\infty e^{x\omega}\Phi_{s-1}(x)\,dx = (-\omega)^2 \int_0^\infty e^{x\omega}\Phi_{s-2}(x)\,dx$$

$$(-\omega(^2 \int_0^\infty e^{x\omega}\Phi_{s-2}(x)\,dx = (-\omega)^3 \int_0^\infty e^{x\omega}\Phi_{s-3}(x)\,dx$$

. . . . . .

. . . . . .

$$(-\omega)^{s-1} \int_0^\infty e^{x\omega}\Phi_1(x)\,dx = (-\omega)^s \int_0^\infty e^{x\omega}\Phi_0(x)\,dx.$$

Or finally

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx = (-\omega)^s \int_0^\infty e^{x\omega}\Phi_0(x)\,dx.$$

Expanding $e^{x\omega}$ in a power series we have

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx =$$

$$= \frac{(-\omega)^s}{n\sqrt{2\pi}} \int_0^\infty \left[1 + x\omega + \frac{x^2\omega^2}{2!} + \frac{x^3\omega^3}{3!} + \ldots\right] e^{-\frac{1}{2}\left[\frac{\log x - m}{n}\right]^2} dx.$$

The general term in this expansion is of the form

$$\frac{(-\omega)^s}{n\sqrt{2\pi}} \frac{\omega^r}{r!} \int_0^\infty x^r e^{-\frac{1}{2}\left[\frac{\log x - m}{n}\right]^2} dx$$

which according to the formulas given on page 74 reduces to:

$$(-\omega)^s e^{m(r+1)+\frac{1}{2}m^2(r+1)^2}\omega^r : r!$$

Hence we may write

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx = (-\omega)^s \sum_{r=0}^{r=\infty} e^{m(r+)1+\frac{1}{2}m^2(r+1)^2}\omega^r : r!$$

Consequently the relation between the semi-invariants and the frequency function

$$F(x) = k_0\Phi_0(x) - \frac{k_1}{1!}\Phi_1(x) + \frac{k^2}{2!}\Phi_2(x) - \frac{k_3}{3!}\Phi_3(x) + \ldots$$

can be expressed by the following recursion formula

$$s_0 e^{\frac{\lambda_1\omega}{1!}+\frac{\lambda_2\omega^2}{2!}+\frac{\lambda_3\omega^3}{3!}+\cdots} = s_0 + \frac{s_1\omega}{1!} + \frac{s_2\omega^2}{2!} + \frac{s_3\omega^3}{3!} + \ldots =$$

$$= \sum_{v=0}^{v=\infty} s_v \frac{\omega^r}{v!} = \sum_{s=0}^{s=\infty} \frac{k_s}{s!}\omega^s \sum_{r=0}^{r=\infty} e^{m(r+1)+\frac{1}{2}m^2(r+1)^2}\omega^r : r!$$

The constants k are here expressed in terms of the unadjusted moments or power sums, s. It is readily seen that the Sheppard corrections for adjusted moments, M, also apply in this case. We are, therefore, able to write down the values

of the k's from the above recursion formula in the following manner

$$M_0 = k_0 e^{m+1/2n^2}$$
$$M_1 = k_1 e^{m+1/2n^2} + k_0 e^{2m+2n^2}$$
$$M_2 = k_2 e^{m+1/2n^2} + 2k_1 e^{2m+2n^2} + k_0 e^{3m+4.5n^2}$$
$$M_3 = k_3 e^{m+1/2n^2} + 3k_2 e^{2m+2n^2} + 3k_1 e^{3m+4.5n^2} + k_0 e^{4m+8n^2}$$
$$M_4 = k_4 e^{m+1/2n^2} + 4k_3 e^{2m+2n^2} + 6k_2 e^{3m+4.5n^2} + 4k_1 e^{4m+8n^2} + k_0 e^{5m+12.5n^2}$$

It is easy to see that it is not possible to determine the generating function's parameters m and n from the observations. These parameters like M and σ in the case of the Laplacean normal probability curve must be chosen arbitrarily. If m and n are selected so as to make k_1 and k_2 vanish we have

$$M_0 = k_0 e^{m+1/2n^2}$$
$$M_1 = k_0 e^{2m+2n^2}$$
$$M_2 = k_0 e^{3m+4.5n^2}$$

the solution of which gives

$$e^{n^2} = \frac{M_0 M_2}{M_1^2}, \quad e^{2m} = \frac{M_1^3}{M_0^2 M_2}, \quad k_0 = \frac{M_0^3 M_2}{M_1^3}$$

while

$$k_3 e^{m+1/2n^2} = M_3 - M_0 e^{3m+7.5n^2}$$

$$k_4 e^{m+1/2n^2} = M_4 - 4M_3 e^{m+1.5n^2} - M_0 e^{4m+9n^2}(e^{3n^2} - 4).$$

This theory requires the computation of a set of tables of the generating function

$$\Phi_0(x) = \frac{1}{n\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\log x - m}{n}\right]^2}$$

and its derivatives. For $\Phi_0(x)$ itself we may of course use the ordinary tables for the normal curve $\varphi_0(z)$ when we consider

$$z = \frac{\log x - m}{n}.$$

I have calculated a set of tables of the derivatives of $\Phi_0(x)$ and hope to be able to publish the manuscript thereof in the second volume of my treatise on "*The Mathematical Theory of Probabilities*".

23. PARAMETERS DETERMINED BY LEAST SQUARES

The above development is based upon the theory of functions and the theory of definite integrals. We shall now see how the same problem may be attacked by the method of least squares after we have determined by the usual method of moments the values of m and n in the generating function $\varphi_0(z)$.

Viewed from this point of vantage our problem may be stated as follows:

Given an arbitrary frequency distribution, of the variate z with $z = (\log x - m) : n$ and where x is reckoned from a zero point or origin, which is situated a units below the mean and defined by the relation

$$\eta^3\lambda_3 - 3\eta^2\lambda_2^2 = \lambda_2^3, \text{ where } a = \lambda_1 - \eta;$$

to develop $F(z)$ into a frequency series of the form

$$F(z) = k_0\varphi_0(z) + k_3\varphi_3(z) + k_4\varphi_4(z) + \ldots + k_n\varphi_n(z),$$

where the k's must be determined in such a way that the expression

$$\sum_{i=0}^{i=n} k_i\varphi_i(z)$$

gives the best approximation to $F(z)$ in the sense of the method of least squares.

Stated in this form the frequency function is reduced to the ordinary series of Gram or the A type of the Charlier series, already treated in the earlier chapters.

24. APPLICATION TO GRADUATION OF A MORTALITY TABLE

As an illustration of the theory to a practical problem we present the following frequency distribution by 5-year age intervals of the number of deaths (or Σd_x by quinquennial grouping) in the recently published American-Canadian Mortality of Healthy Males, based on a radix of 100,000 entrants at age 15.

Frequency Distribution of Deaths by Attained Ages in American-Canadian Mortality Table.

Ages	Σdx	1st Component	2d Comp.
15— 19	1,801	120	1,681
20— 24	1,996	230	1,766
25— 29	2,089	440	1,649
30— 34	2,120	790	1,330
35— 39	2,341	1,370	971
40— 44	2,911	2,270	641
45— 49	3,937	3,570	367
50— 54	5,527	5,400	127
55— 59	7,723	7,722	1
50— 64	10,383	10,383	
65— 69	12,987	12,987	
70— 74	14,535	14,535	
75— 79	13,807	13,807	
80— 84	10,328	10,328	
85— 89	5,464	5,464	
90— 94	1,757	1,757	
95— 99	278	278	
100—104	16	16	
	100,000	91,467	8,533

The curve represented by the d_x column is evidently a composite frequency function compounded of several series. From a purely mathematical point of view the compound curve may be considered as being generated in an infinite number of ways as the summation of separate component frequency curves. From the point of view of a practical graduation it is, however, easy to break this compound death curve up into two separate components. A mere glance at the d_x curve itself suggests a major skew frequency curve with a maximum point somewhere in the age interval from 70—75 and minor curve (practically one-sided) for the younger ages.

Let us therefore break the Σd_x column up into the two so far perfectly arbitrary parts as shown in the above table and then try to fit those two distributions to logarithmically transformed A curves.

Starting with the first component the straightforward computation of the semi-invariants is given in the table below with the provisional mean chosen at age 67.

Frequency Distribution of Deaths in American Mortality Table First Component.

Ages	x	$F(x)$	$xF(x)$	$x^2F(x)$	$x^3F(x)$
104—100	— 7	16	112	784	5,488
99— 95	— 6	278	1,668	10,008	60,048
94— 90	— 5	1,757	8,785	43,925	219,625
89— 85	— 4	5,464	21,856	87,424	349,696
84— 80	— 3	10,328	30,984	92,952	278,856
79— 75	— 2	13,807	27,614	55,228	110,456
74— 70	— 1	14,535	14,535	14,535	14,535
69— 65	— 0	12,987	0	0	0
		59,172	105,554	304,856	1,038,704
64— 60	+ 1	10,383	10,383	10,383	10,383
59— 55	+ 2	7,723	15,446	30,892	61,784
54— 50	+ 3	5,400	16,200	48,600	145,800
49— 45	+ 4	3,570	14,280	57,120	228,480
44— 40	+ 5	2,270	11,350	56,750	283,750
39— 35	+ 6	1,370	8,220	49,320	295,920
34— 30	+ 7	790	5,530	38,710	270,970
29— 25	+ 8	440	3,520	28,160	225,280
24— 20	+ 9	230	2,070	18,630	167,670
19— 15	+10	120	1,200	12,000	120,000
		32,296	88,199	350,565	1,810,037
	Sr	91,468	—17,355	655,421	771,333

Computing the semi-invariants by means of the usual formulas in paragraph 13, we have:

$\lambda_1 = -17355 : 91468 = -0.18974$, or mean at age 67 + 5 (0.19) or at age 67.95

$\lambda_2 = 655421 : 91468 - \lambda_1^2 = 7.1296$
$\lambda_3 = 771333 : 91468 - 3\lambda_1 m_2 + 2\lambda_1^3 = 12.4981.$

In order to determine the mathematical zero or the origin we have to solve the following cubic:

$$\lambda_3 \eta^3 - 3\lambda_2^2 \eta^2 = \lambda_2^3, \text{ or}$$
$$12.498\,\eta^3 - 152.511\,\eta^2 = 362.47$$

the positive root of which is equal to 12.39. The zero point is therefore found to be situated 12.39 5-year units from the mean or at age 67.95 + 5 (12.39), i. e. very nearly at age 130, which we henceforth shall select as the origin of the co-ordinate system of the first component. We have furthermore

$$12.39 = e^{m + 1.5n^2}, \text{ and } 7.1296 = e^{2m + 3n^2}(e^{n^2} - 1) =$$
$$= (12.39)^2 (e^{n^2} - 1),$$

the solution of which gives $n^2 = 0.04436$, $n = 0.2106$, $m = 2.4504$, all on the basis of a 5-year interval as unit. If we wish to change to a single calendar year unit we must add the natural logarithm of 5, or 1.6094, to the above value of m, which gives us $m = 4.0598$, while n remains the same. The above computations furnish us with the necessary material for the logarithmic transformation of the variate x which now may be written as

$$z = [\log\ (130 - x) - 4.0598] : 0.2106,$$

where x is the original variate or the age at death.

Having thus accomplished the logarithmic transformation we may henceforth write the generating function as

$$\Phi_0(x) = \frac{1}{.2106\sqrt{2\pi}}\, e^{\frac{1}{2}\left[\frac{\log(130-x) - 4.0598}{0.2106}\right]^2} =$$

$$= \varphi_0(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2:2}$$

We express now $F(x)$ by the following equation.

$$F(x) = k_0\Phi_0(x) + k_3\Phi_3(x) + k_4\Phi_4(x) + \ldots.$$

or in terms of the transformed z:

$$\varphi(z) = k_0\varphi_0(z) + k_3\varphi_3(z) + k_4\varphi_4(z) + \ldots.,$$

and proceed to determine the numerical values of k by the method of least squares.

The numerical calculation required by this method follows precisely along the same lines as described in paragraph 17. I shall for this reason not reproduce these calculations but limit myself to quote the final results for the various coefficients k, which are as follows :—[1]

[1] Interested readers may consult the detailed computations on pages 246—257 in my Mathematical Theory of Probabilities (2nd Edition, New York, 1921.

$$k_0 = 7361.8; \quad k_3 = -212.2; \quad k_4 = -9.6.$$

The final equation of the frequency curve of the first component $F(x)$, is therefore :—

$$F_I(x) = 7361.8\varphi_0(z) - 212.2\varphi_3(z) - 9.6\varphi_4(z),$$

where the generating function, $\varphi_0(z)$, is of the form :—

$$\varphi_0(z) = \Phi_0(x) = \frac{1}{0.2106\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\log(130-x) - 4.0598}{0.2106}\right]^2}$$

The second component, $F_{II}(x)$, can by means of a similar process be expressed by the equation :—

$$F_{II}(x) = 947.4\varphi_0(z) - 63.4\varphi_3(z) - 30.0\varphi_4(z),$$

where

$$\varphi_0(z) = \Phi_0(x) = \frac{1}{0.12\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\log(x+68.8) - 4.532}{0.12}\right]^2}.$$

Addition of these two component curves gives us the ultimate compound frequently curve, representing the d_x of the mortality table.

A comparison between the observed values of d_x and the values of d_x as computed from the above equation is shown in graphical form in the attached diagram. Evidently the graduation leaves but little to be desired in the way of closeness of fit.

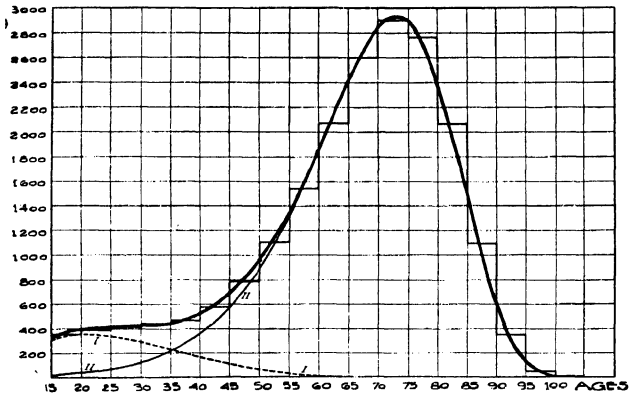

Figure 1.

Diagram showing graduation of d_x column in the *AM* (5) table by a compound frequency curve of the Gram-Charlier types.

25. BIOLOGICAL INTERPRETATION OF MORTALITY

It appears that the Italian statisticians were the first to break up the d_x curve into a system of five or more component frequency curves, which, however, were all of the normal Laplacean type. Pearson who in a brillant essay entitled *Chances of Death* was the next to attack the problem, employed a system of five skew frequency curves. Already as early as 1914 I found that from ages above 10 the majority of d_x curves in previously constructed mortality tables could be represented by not more than two skew

frequency curves as shown in the above example of the *AM* (5) table.

Although all such investigations may be very interesting and useful from the point of view of the actuary, we must, however, not overlook the fact that the breaking up of the compound d_x curve in the manner just described is merely an empirical process pure and simple. While such processes undoubtedly represent very neat methods of graduation, a quite different and more important question is whether mathematical work of this kind allows of a biological interpretation. It is evident that from a mere mathematical point of view we may break up the d_x curve into various component parts in an infinite number of ways. But while such breaking up processes may be extremely interesting as actuarial graduations and exercises in pure mathematics, they have evidently little connection with the underlying biological facts of a mortality table. This aspect of the question has been brought out in a very forcible manner by the eminent American biometrician, Raymond Pearl, in his 1920 Lowell Institute Lectures. The whole subject would appear in a quite different light if it were possible to give a biological interpretation of the mathematical analysis and to show that the component frequency curves as derived from pure mathematics

have a counterpart in actual life. This, I think, would be very difficult, if not impossible to establish, because it is not mathematics which determines the conduct or behavior of living organisms. One might, however, view the whole problem from the standpoint of the biologist rather than from the standpoint of the mathematician. The problem then is to ascertain whether the observed biological facts as shown in the collected statistical data allow of a mathematical interpretation, rather than to find a biological interpretation and counterpart of previously established empirical formulae.

It is to this important question that I have devoted the entire discussion of the second chapter of this book. I have proceeded from certain observed biological facts (in this particular instance the statistics on the number of deaths by sex and attained ages from more than 150 causes of death) which represent the natural phenomena under investigation. In order to offer a rational explanation of these facts and to interprete their quantitative relationships, I have adopted as a working hypothesis the supposition that the number of deaths according to attained age and sex among the survivors of a homogeneous cohort of say 1,000,000 entrants at age 10 tend to cluster around specific ages in such a manner

that their frequency distribution by attained ages can be represented by a limited number of sets of Gram-Charlier or Poisson-Charlier frequency curves.

On the basis of this hypothesis we can now by simple mathematical deductions construct a mortality table from deaths by sex, age and cause of death and without any information about the lives exposed to risk at various ages.

Finally we can verify the ultimate results contained in this final mortality table by working back from the table to the data originally observed.

This procedure is in strict conformity with the model of modern science, which according to Jevons consists of the four processes of *observation*, *hypothesis*, *deduction* and *verification*.

The important factor in this investigation, and one which most actuaries and statisticians fail to grasp, is that I have looked at the whole problem as a biometrician rather than as a mathematician. Mathematics has been employed only as a working tool in the whole process, and the reason that the method has met with success must be sought for in concrete biological facts and not in the realm of mathematics.

26. POISSON'S PROBABILITY FUNCTION

In certain statistical series it frequently happens that the semi-invariants of higher order than zero all are equal, or that

$$\lambda_1 = \lambda_2 = \lambda_3 = \ldots = \lambda_r = \lambda.$$

We shall for the present limit our discussion to homograde statistical series where the variates always are positive and integral, and where therefore the definition of the semi-invariants is of the form :—

$$e^{\frac{\lambda\omega}{1!}+\frac{\lambda\omega^2}{2!}+\frac{\lambda\omega^3}{3!}+\ldots} \Sigma\varphi(x) = \Sigma\varphi(x)e^{x\omega} =$$

$$= \varphi(0)\,e^{0\omega} + \varphi(1)e^{1\omega} + \varphi(2)e^{2\omega} + \varphi(3)e^{3\omega} + \ldots,$$

or

$$e^{\frac{\lambda\omega}{1!}+\frac{\lambda\omega^2}{2!}+\frac{\lambda\omega^3}{3!}+\ldots} = e^{-\lambda}e^{\lambda e^{\omega}} = \Sigma\varphi(x)e^{x\omega}$$

for $x = 0, 1, 2, 3, \ldots,$

which also can be written as

$$e^{-\lambda}\left(1+\frac{\lambda e^{\omega}}{1!}+\frac{\lambda^2 e^{2\omega}}{2!}+\ldots.\right) =$$

$$= \varphi(0)1 + \varphi(1)e^{\omega} + \varphi(2)e^{2\omega} + \ldots.$$

The coefficient of $e^{r\omega}$ gives the relative frequency or the probabitity for the occurence of $x = r$, and we find therefore that

$$\varphi(x) = \psi(r) = \frac{e^{-\lambda}\lambda^r}{r!}.$$

This is the famous Poisson Exponential, so called after the French mathematician, Poisson, who first derived this expression in his *Recherches sur la Probabilites des jugesments*, but in an entirely different manner than the one we have indicated above.

The Poisson Exponential opens a new way for the treatment of statistical series which possess the attribute that all their semi-invariants of higher order than zero are all equal, or nearly equal. It is readily seen that whereas the Laplacea probability function $\varphi_0(x)$ contains two parameters λ_1 and σ the probability function of Poisson contains only one parameter, λ.

27. POISSON—CHARLIER FREQUENCY CURVES

We have already seen in the previous chapters that the Gram-Charlier frequency curve could be written as

$$F(x) = \Sigma c_i \varphi_i(x) = \Sigma c_i H_i(x) \varphi_0(x)$$
$$\text{for } i = 0, 1, 2, 3, \ldots .$$

where $\varphi_0(x)$ is the generating Laplacean probability function.

The idea now immediately suggests itself to

use a similar method of expansion in the case of the Poisson probability function and to employ this exponential as a generating fuction in the same manner as the Laplacean function. We are, however, in the present case of the Poisson exponential dealing with a generating function which so far has been defined for positive integral values only and, therefore, represents a discrete function. For this reason it will be impossible to express the series as the sum-products of the successive derivatives of the generating function and their correlated parameters c. We can, however, in the case of integral variates express the series by means of finite differences and write $F(x)$ as follows:

$$F(x) = c_0\psi(x) + c_1\Delta\psi(x) + c_2\Delta^2\psi(x)\ldots. \quad (I)$$

where $\psi(x) = e^{-m}m^x{:}x!$ for $x = 0, 1, 2, 3, \ldots.$, and

$$\Sigma\psi(x) = 1,$$
$$\Delta\psi(x) = \psi(x) - \psi(x-1),$$
$$\Delta^2\psi(x) = \Delta\psi(x) - \Delta\psi(x-1) = \psi(x) - 2\psi(x-1) + \psi(x-2).$$

The series (I) is known as the Poisson-Charlier frequency series or Charlier's B type of frequency curves.

The semi-invariants of these frequency series are given by the following relation:

$$e^{\frac{\lambda_1\omega+\lambda_2\omega^2+\lambda_3\omega^3+\ldots}{2!\quad 3!}} =$$

$$= \sum_{x=-\infty}^{x=0} [c_0\psi(x) + c_1\Delta\psi(x) + c_2\Delta^2\psi(x) + \ldots]\, e^{x\omega}.$$

Expanding and equating the co-efficients of equal powers of ω we have:

$$\lambda_0 = 1 = c_0\Sigma\psi(x) \text{ or } c_0 = 1$$
$$\lambda_1 = \Sigma x\,(\psi(x) + c_1\Delta\psi(x) + c_2\Delta^2\psi(x) + \ldots) \quad (II)$$
$$\lambda_1^2 + \lambda_2 = \Sigma x^2(\psi(x) + c_1\Delta\psi(x) + c_2\Delta^2\psi(x) + \ldots)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$

We now have

$$\Sigma\psi(x) = 1, \text{ and}$$
$$\Sigma x\psi(x) = \Sigma m e^{-m} m^{x-1} : (x-1)! = m\Sigma\psi(x-1) = m.$$

We also find from well-known formulas of the calculus of finite differences that[1]

$$\Sigma x^2\psi(x) = m^2 + m$$
$$\Sigma x\Delta\psi(x) = -1$$

[1] These formulas can also be derived from the definition of the semi-invariants and the well-known relations between moments and semi-invariants as given on page 74 when we remember that according to our definition all semi-invariants in the Poisson exponential are equal to m.

$$\Sigma x\Delta^2\psi(x) = 0$$
$$\Sigma x^2\Delta\psi(x) = -(2m + 1)$$
$$\Sigma x^2\Delta^2\psi(x) = 2$$

Substituting these values in (*II*) we obtain

$$\lambda_1 = m - c_1$$
$$\lambda_1^2 + \lambda_2 = m^2 + m - (2m + 1)\ c_1 + 2c_2$$

By letting $m = \lambda_1$ we can make the coefficient c_1 vanish, which results in

$$\lambda_1 = m$$
$$c_2 = \tfrac{1}{2}[\lambda_2 - m]$$

where the two semi-invariants λ_1 and λ_2 are calculated around the natural zero of the number scale as origin.

For the above discussion we have limited ourselves to the determination of the three constants m, c_0 and c_2. It is easy, however, to find the higher parameters c_3, c_4, c_5, . . . from the relations between the moments of the Poisson function and the semi-invariants of order 3, 4, 5, . . . ect. Charlier usually calls the parameter m the *modulus* and c_2 the *eccentricity* of the B curve.

28. NUMERICAL EXAMPLES

As an illustration of the application of the Poisson-Charlier series we select the following series of observations on alpha particles radiated from a bar of Polonium as determined by Rutherford and Geiger.

The appended table states the number of times, $F(x)$, the number of particles given off in a long series of intervals, each lasting one-eighth of a minute had a given value x: —

x	$F(x)$	x	$F(x)$	x	$F(x)$
0	57	5	408	10	10
1	203	6	273	11	4
2	383	7	139	12	0
3	525	8	45	13	1
4	532	9	27	14	1

We are here dealing with integral variates which can assume positive values only and the observations are therefore eminently adaptable to the treatment by Poisson-Charlier curves. Selecting the natural zero as the origin of the coordinate system we find that the first two semi-invariants are of the form

$\lambda_1 = 3.8754$, $\lambda_2 = 3.6257$, and we therefore have: $m = \lambda_1 \quad 3.88$; $c_2 = \frac{1}{2}[\lambda_2 - m] = -0.125$.

The equation for the frequency distribution of the total $N = 2608$ elements therefore becomes

$$F(x) = N[\psi_{3\cdot88}(x) + (-0.125)^2\Delta\psi_{3\cdot88}(x)].$$

The table below gives the values as fitted to the curve, $F(x)$:

Alpha Particles Discharged from Film of Polonium (Rutherford and Geiger).

$N = 2608$, $m = 3.88$, $c_2 = -0.125$

(1) x	(2) $\psi(x)$	(3) $\Delta^2\psi(x)$	(4) $N\times(2)$	(5) $N\times(3)\times c_2$	(6) (4) + (5)
0	.020668	+.020668	53.9	— 6.7	47
1	.080156	+.038820	209.0	—12.7	196
2	.155455	+.015811	405.4	— 5.2	400
3	.201015	—.029793	524.2	+ 9.7	533
4	.194967	—.051608	508.5	+16.8	525
5	.151625	—.037654	394.5	+12.3	407
6	.097850	—.009714	254.9	+ 3.2	258
7	.054249	+.009814	141.2	— 3.2	138
8	.026316	+.015668	68.7	— 5.1	64
9	.011351	+.012968	29.6	— 4.2	25
10	.004407	+.008021	11.5	— 2.6	9
11	.001555	+.004092	4.1	— 1.2	3
12	.000503	+.001800	1.3	— 0.6	1
13	.000150	+.000699	0.4	— 0.2	0
14	.000042	+.000245	0.1	— 0.1	0
15	.000010	+.000076	0.0	— 0.0	0
16	.000003	+.000025	——	——	0
17	.000001	+.000005	——	——	0

As a second example we offer our old friend, the distribution of flower petals in *Ranunculus Bulbosus*. Selecting the zero point at $x = 5$ and

computing the semi-invariants in the usual manner we obtain the following equation for the frequency curve.

$$F(x) = 222\,\psi(x) + 31.5\,\Delta^2\psi(x), \quad m = 0.631.$$

A comparison between calculated and observed values follows :—

x	$F(x)$	Obs.
5	134.9	133
6	51.6	55
7	22.5	23
8	9.5	7
9	2.9	2
10	0.6	2

29. TRANSFORMATION OF THE VARIATE

For integral variates we have shown that the Poisson frequency curve possesses the important property that all its semi-invariants are equal. Now while a frequency distribution of a certain integral variate, x, may perhaps *not* possess this property, it may, however, very well happen after a suitable linear transformation has been made, that the variate thus transformed will be subject to the laws of Poisson's function.

Let $z = ax - b$ represent the linear transformation which is subject to the above laws with a series of semi-invariants all equal to m.

These semi-invariants according to the properties set forth in paragraph 5 are therefore

$$\begin{aligned} m &= \lambda_1(z) = a\lambda_1(x) - b \\ m &= \lambda_2(z) = a^2\lambda_2(x) \\ m &= \lambda_3(z) = a^3\lambda_3(x) \\ &\quad . \quad . \quad . \quad . \quad . \quad . \quad . \end{aligned}$$

and our problem is to find the unknown parameters a, b and m.

Simple algebraic methods, which it will not be necessary to dwell upon, give the following results:

$$\begin{aligned} a &= \lambda_2 : \lambda_3 \\ m &= \lambda_2^3 : \lambda_3^2 \\ b &= a\lambda_1 - m \end{aligned}$$

As a numerical illustration of this transformation we choose from Jørgensen a series of observations by Davenport on the frequency distribution of glands in the right foreleg of 2000 female swine.

No. of Glands ..	0	1	2	3	4	5	6	7	8	9	10
Frequency	15	209	365	482	414	277	134	72	22	8	2

The values of the three first semi-invariants are

$\lambda_1 = 3.501$, $\lambda_2 = 2.825$, $\lambda_3 = 2.417$,

$a = 2.825 : 2.417 = 1.168$,

$m = 2.825^3 : 2.417^2 = 3.859$,

$b = (1.168)\ (3.501) - 3.859 = 0.230$.

The new variable then becomes $z = az - b$ and the transformed Poisson probablity function takes on the form:

$$\psi(z) = \frac{e^{-m} m^z}{z!}.$$

In general, however, we will find that z is not a whole number and the expression $z!$ therefore has no meaning from the point of view of factorials at least. This difficulty may, however, be overcome through the introduction of the well-known Gamma Function, $\Gamma(z+1)$, which holds true for any positive or negative real value of z and which in the case of integral values of z reduces to $\Gamma(z+1) = z!$

Hence we can write the transformed Poisson probability function as

$$\psi(z) = \frac{e^{-m} m^z}{\Gamma(z+1)}.$$

Tables to 7 decimal places of the Gamma Function, or rather for the expression $-\Gamma(z+1)$, have been computed by Jørgensen in his Frekvens-

flader and Korrelation from $z = -5$ to $z = 15$, progressing by intervals of 0.01.

By means of this table and the tables of ordinary logarithms it is now easy to find the values of $\psi(z)$ in the case of the example relating to the number of glands in female swine. The detailed computation is shown below.[1]

(1)	(2)	(3)	(4)	(5)	(6)	(7)
x	z	$-\log \Gamma(z+1)$	$\log m^z$	$(3)+(4)$ $+\log e^{-m}$	$\psi(z)$	$F(x)$
0	—.230	.9209	.8651	.1101--2	.0129	30.1
1	+.938	.0108	.5500	.8849--2	.0767	179.2
2	2.106	.6555	.2350	.2146—1	.1639	382.9
3	3.274	.0679	.9199	.3119—1	.2051	479.1
4	4.442	.3216	.6048	.2501--1	.1780	415.8
5	5.610	.4547	.2897	.0685--1	.1171	273.6
6	6.778	.4904	.9746	.7891--2	.0615	143.7
7	7.946	.4446	.6595	.4282—2	.0268	62.6
8	9.114	.3285	.3444	.9970—3	.0099	23.1
9	10.282	.1506	.0294	.5041—3	.0032	7.5
10	11.450	.9177	.7143	.9561--4	.0009	2.1

[1] The characteristics of the logarithms have been omitted in this table (except in column 5) and only the positive mantissas are shown. Column 7 represents the 2000 individual observations pro rated according to column 6.

CHAPTER II

(TRANSLATED BY MR. VIGFUSSON)

THE HUMAN DEATH CURVE

1. INTRODUCTORY REMARKS

In the following paragraphs I intend to discuss a method of constructing mortality tables from mortuary records by sex, age and cause of death, but without reference to or knowledge of the exposed to risk at various ages. This proposed method is indeed one which has been severely criticized in certain quarters, and several critics flatly deny that it is possible to construct mortality tables from such data without detailed information of the exposed to risk. It is, however, a very dangerous practice to say that a certain thing is impossible. The true scientist, least of all, should attempt to set limits for the extension of human knowledge. It is still remembered how the great August Comte once denied that it ever would be possible to determine the chemical constituents of the celestial bodies. Only a few years after this emphatic denial by the brilliant French-

man the spectroscope was discovered, by means of which we have been able to detect a number of chemical elements of other worlds than that of our own little earth. It is but fair to say that the method which we here shall describe has met with rather determined opposition in certain actuarial quarters. Under such circumstances it is natural that the process will be viewed in a light of scepticism and criticism. I welcome such an attitude because it has been my purpose to present the following studies for further investigation and not to force them upon my readers as authoritative or as a kind of infallible dogma.

In presenting the outlines of the proposed method I wish to state that it has never been the intention to supplant the orthodox methods of constructing mortality tables where we have exact information of the so-called "exposed to risk" or number living at various ages. Numerous and very important examples, however, offer themselves in actuarial and statistical practice where such information is not available. Most of the greater American Life Insurance Companies, especially those writing the so-called industrial insurance, have on hand an enormous amount of information of deaths by sex, attained age and by cause of death among their policyholders. Even the mortuary records of certain occupations, as for instance metal and coal miners, among the

death claims in the industial class are so numerous, that it would be possible to construct a mortality table for such professions if we know the exact number exposed to risk at various ages. Such information is, however, in the majority of cases wanting, or could only be obtained by means of a great expenditure of time and labor. Again, as Mr. F. S. Crum has pointed out in an article in the "Insurance and Commercial Magazine", a number of cities and states in United States give from year to year very detailed information in regard to mortuary records by sex, age at death and cause of death. On account of the intense migration taking place in certain sections of the United States, especially in those of an industrial character, it is, however, impossible to know the exact population at various ages, except in the particular years in which the federal or state census has been taken. The fact that for all but a few states of this country the intercensal period is no less than ten years, the determination of the population composition by age and sex for a given locality and intercensal year, with any degree of accuracy, becomes a practical impossibility without a special count. Such a count or census of a specific locality or a single city is, however, a costly undertaking at its best, for which the necessary funds are rarely available. In all such instances the mortuary records are practically

worthless in so far as the construction and computation of death rates are concerned, if we are to rely solely upon the usual method of constructing mortality tables. It will therefore readily be seen that, apart from purely academic interests, the possibility of establishing a method of constructing mortality tables without knowing the population exposed to risk at various ages would be of great practical value, and I deem no apology necessary to present the following method, which intends to overcome this very obstacle of having no information of the exposures.

2. EMPIRICAL AND INDUCTIVE METHODS OF SOLUTION.

In order to bring the method into the proper perspective it will be of value to contrast it with the ordinary methods followed in the construction of mortality tables. Let us therefore briefly review those methods and principles commonly employed by actuaries and statisticians. A certain number, say L_x persons at age x, are kept under observation for a full calendar year and the number, D_x, who die among the original entrants during the same year are recorded. The ratio $D_x : L_x$ is then considered as the crude probability of dying at age x. Similar crude rates are obtained for all other ages and are then subjected to a more or less empirical process of graduation to

smooth out the irregularities arising from what is considered as random sampling. One then chooses an arbitrary radix, say for instance 100,000 persons at age 10, which represents a hypothetical cohort of 10-year old children entering under our observation. This radix is then multiplied by the previously constructed value of q_{10} and the product represents the number dying at age 10. This number, d_{10}, is subtracted from l_{10} or 100,000 and the difference is the number living at age 11 or l_{11}. This latter number is then multiplied by q_{11} and the result is d_{11}, or the number dying at age 11 out of the original cohort of 100,000. In this way one continues for all ages up to 105, or so.

It is to be noted that the column of q_x in this process represents the fundamental column while the columns of l_x and d_x are purely auxiliary columns.

Allow us here to ask a simple question. Do these empirically derived numbers of deaths at various ages out of an original cohort of 100,000 entrants at age 10 give us any insight or clue as to the exact nature of the biological phenomenon known as death, and are we by this method enabled to lift the veil and trace the numerous causes which must have been at work and served to produce the total effect, the d_x curve, of which we by means of the usual methods have a purely

empirical representation? I fear that this question will have to be answered in the negative. The usual actuarial methods do not give us a single glance into the relation between cause and effect, which after all is the ultimate object of investigation for all real science. Probably some critics would answer that they are not interested in investigating causal relations. Such an attitude of indifference is, however, very dangerous for a statistician or an actuary whose very work rests upon the validity of the law of causality. We may, however, overlook this apparent inconsistency of the empiricists and turn our attention to the proposed methods of constructing mortality tables along inductive lines, or by the process which Jevons has termed a complete induction.

Such a process we should find diametrically opposite to the methods of the empiricists, both in respect to points of attack and deduction. In the case of the empiricists the q_x is the initial and fundamental function from which the d_x column is computed as a mere by-product. The rationalistic method starts with the d_x column and terminates with the q_x as the by-product.

Being primarily interested in the absolute number of deaths and not in the relative frequencies of deaths at various ages, our first question is therefore, "What is the form of the frequency

curve representing the deaths at various ages among the survivors of the original group of 100,000 entrants at age 10?" Right here we can, strange to say, apply some purely a priori knowledge. We know a priori that the curve must be finite in extent, because of the very fact that there is a definite limit to human life, and we also know that it assumes only positive values. There can be no negative numbers of deaths unless we were to regard the reported theological miracles of resurrections from the Jewish-Christian religion as such. This information about the death curve, or the curve of d_x, is, however, not sufficient for use as a basis for our deductions. We must therefore look about for additional information, whether of an *a priori* or an *a posteriori nature* and of such *a general character that it can be adopted as a hypothesis.*

3. GENERAL PROPERTIES OF THE "DEATH CURVE"

It was Poincare who once said that every generalization is a hypothesis. Hence we shall look for some *general* characteristics which *all* mortality tables have in common in the age interval under consideration (age 10 and upwards). Let us take any mortality table, I do not care from what part of the world, and examine the general trend of the curve traced

by the values of d_x for various ages. The curve rises gradually from the age of ten. The increase in the number of deaths among the survivors at various ages will increase, although not uniformly, until the ages around 70 or 75 are reached. At this age interval we generally encounter a maximum. From the ages between 70 and 75 and for higher ages the number of deaths among the survivors will decrease at a more rapid rate than at the earlier stages of life. After the age of 85 only a small number of the veteran cohort are still alive. After the age of 90 only a few centenarians struggle along, keeping up a hopeless fight with the grim reaper, Death, until eventually all are carried off between the ages of 110 and 115. We can much better illustrate this process of the struggle between the surviving members at various ages of the cohort and the opposing forces as marshalled by the ultimate victor, Death, through a graphical representation. The chart on page 114 shows a mortality graph of the male population in Denmark (1906-1910) from ages 10 and upwards as constructed by the Royal Danish Statistical Bureau. The ordinates of the curve show the number of deaths at various ages among the survivors of the original cohort of 100,000 entrants at age10. We notice a gradual increase from the younger ages until the age of 77, where a max-

imum or high crest is encountered. From that age a rapid decline takes place until the curve approaches the abscissa with a strongly marked asymptotic tendency after the age of 90. At the age of 110 all the members of the cohort have lost out and death stands as the undisputed victor, a victor among a mass of graves. The curve we thus have traced may properly be called "The Curve of Death". On the same chart I have also shown a graphical representation of a comparison between the Danish death curve and the corresponding death curves of males for England and Wales in the period 1909—1911, Norway 1900—1910, France 1908--1913 and United States period 1909—1911, all based upon an original radix of 1,000,000 entrants at age 10.

We will notice quite important variations in these curves. The curves for the Scandinavian countries show a relatively heavy clustering around the maximum point which in the case of Denmark is reached at age 75, in England at age 73, and in France at age 72. The Danish curve is also more symmetrical and shows a more uniform clustering tendency around the maximum value than the other curves. The asymmetry or skewness is most pronounced in the American curve, due to the comparatively greater number of deaths at

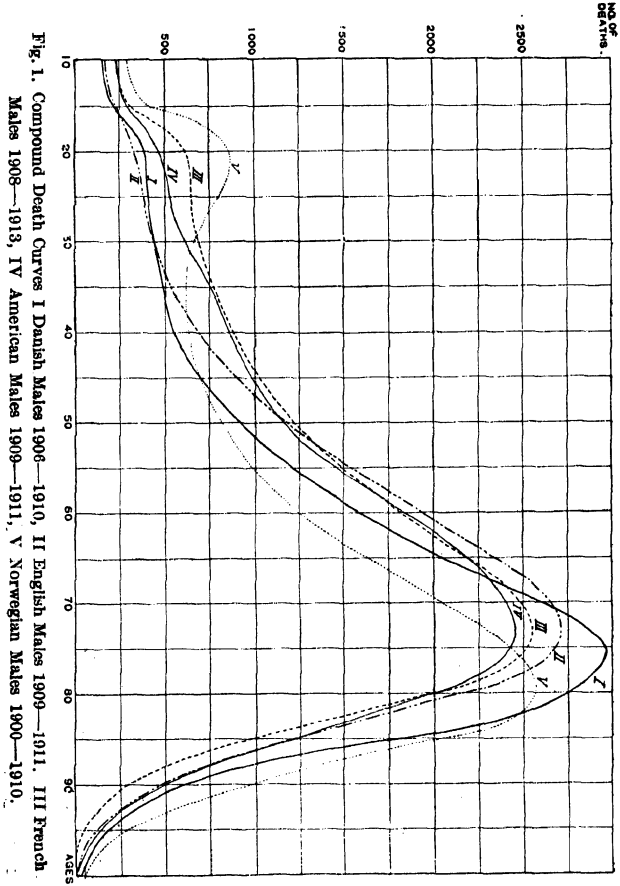

Fig. 1. Compound Death Curves I Danish Males 1906—1910, II English Males 1909—1911, III French Males 1908—1913, IV American Males 1909—1911, V Norwegian Males 1900—1910.

younger ages than in the other tables. In the curve for Norwegian males I might mention

another peculiarity which is absent in most other death curves. I have reference here to a secondary minor maximum or miniature crest at the age of 21. This maximum point, which is not very pronounced arises from the heavy mortality among youths in Norway, whose male population always has consisted of rovers of the sea. A much larger proportion of young men braves the terrors of the sea in Norway than in any country in the world. These sturdy decendents of the Vikings can be found in all parts of the globe. You are sure to find a weatherbeaten Norwegian tramp steamer even in the most deserted and far away harbours of our continents. But the sea takes its toll. The result is shown in the little peak in the curve of death among these sturdy Norwegian youths.[1]

Despite all these smaller irregularities all the curves have, however, certain well defined characteristics, namely:

1) An initial increase with age.
2) A well defined maximum point around the age period 70—80.
2) A more rapid decline from that point until the ultimate end of the mortality table.

[1] Another factor is the high number of deaths from tuberculosis typical of youth. See in this connexion discussion in paragraph 12a under the Japanese Table.

4. RELATION OF FREQUENCY CURVES

The most interesting of these common characteristics is the encountering of a maximum point in the neighborhood of 70, and the subsequent decline toward the higher ages. This fact has a very important biometric significance, which we shall discuss in a somewhat detailed manner. Most of my readers are familiar with the so-called probability curve, expressed by the equation:

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-M)^2/2\sigma^2} \qquad (1)$$

This Laplacean or normal curve is represented in graphical form by the beautiful bellshaped curve so well known to mathematical readers. Various approximations to this curve are continually encountered in numerous instances of observations relating to certain biological phenomena where certain measurable attributes of various sample populations tend to cluster around a certain norm, such as the measurements of heights of recruits, fin rays in fish, etc. We also know that where this tendency to cluster around the mean is asymmetrical or skew, it is in many cases possible to give a very close representation by the Laplacean-Charlier frequency curves.

Now let us return to our curves of death. It

will be noted that all these curves for ages above the crest period 70 to 75 to a very marked degree approach the form of the normal prohability curve and exhibit a marked clustering tendency around this particular period. The ages around 70, the Bible's "three score and ten", can therefore be looked upon as a norm of life around which the deaths of the original cohort group themselves in more or less correspondence with the binomial probability law. This pronounced grouping tendency is a very significant biological phenomenon, which it might be of interest to dwell upon.

If all the members of our original cohort were identical as to physical constitution and characteristics, if they all were exposed to identically the same outward influences acting upon their mode of life, it becomes evident from the law of causality, which is the basis and justification of every collection of statistical data, that all members would die at the same moment. We see, however, immediately that such hypothetical conditions are not present in human society. The paramount feature of our material world is variation. No two persons are alike in regard to physical constitution. Certain inherited characteristics, which are present in the individual in more or less pronounced form, make themselves felt. No two persons or group of persons can be said to be exposed to

the same outward influences. The clergyman and college professor living a sort of tranquil and sheltered life are not exposed to the same dangers as the working man or the man in business life. All these and other factors, almost infinite in number, tend to produce a decided variation in the actual duration of life. Of these influencing factors those relating to purely inherited or natural characteristics are without doubt the most powerful. If it were possible to eliminate certain forms of deaths due to infectious diseases, tuberculosis and accidents, causes more or less due to outward influences, we should have left a number of causes due to a gradual wearing out of the human system, similar in many respects to the deterioration of the mechanism in ordinary machinery. The death curve from such causes of death would be more related to the normal curve than the death curve which includes causes of death from non-inherent or anterior causes as mentioned above. This statement is borne out in the shape of the Danish death curve. In Denmark where a very determined and largely successful fight has been carried on against tuberculosis, and where the accident rate is very low we also find that the curve is more symmetrical than for instance in this country or in England.

This tendency to an approach towards the bi-

nomial probability curve was already noted by Lexis, who from such considerations tried to determine what he called a "Normalalter" or normal age for various countries and sample populations. Speaking of this attempt the eminent Danish statistician, Harald Westergaard, says in his „Statistikens Teori i Grundrids" (Copenhagen 1916) "An unsually interesting attempt has been made by Lexis to determine *the normal age* of man. A mortality table will, as a rule, have two strongly dominant maximum points for the number of deaths. During the first year of life there dies a comparatively large number. From the age of 1 the number of deaths decreases and reaches its lowest point in early youth. It then again begins to increase, at times in wavelike motions, until the maximum point is reached at the old age period".

"The clustering around the latter point has now a great likeness with the normal or Gaussian curve, and we might for this reason call this specific age the *normal life age*. For the calculation of such a normal age the argument may be put forth that experience shows that the great variations in mortality tend to disappear in old age. Let the rate of mortality in a certain generation at age x *be* μ_x and the number of the corresponding survivors be l_x. The quantity $\mu_x l_x$ will

then increase from a certain point, while l_x decreases, in the beginning slowly, but later on at a more rapid pace. "During a long period of life the quantity $\mu_x l_x$—the number of deaths at a certain age—will increase with age. Later on a reversed motion takes place. But when this reversion will occur depends on many conditions, the successful fight against certain diseases, progress in economic conditions, or change in the mode of living. All this exercises an important influence, and the maximum point occurs therefore sometimes sooner and sometimes later. It is also important to investigate the natural selection in old age, which so to say divides the population in different strata, each with its own state of health. The healthiest of such groups will with the increase in age play a greater role. Here as everywhere it is the more important problem to study the clustering around the mean inside the special groups rather than to attempt to find a derived expression for the mortality. On the other hand, the correspondence between the normal curve as established by Lexis is another testimony to the fact that this curve or formula very often can be applied, even in complicated expressions".

5. THE "DEATH CURVE" AS A COMPOUND CURVE

Lexis was satisfied to determine the normal age. A more ambitious attempt to investigate the mortality by means of frequency curves throughout the whole period of life was made by the eminent English biometrician, Pearson, in a brilliant essay in his "Chances of Death". Pearson took the number of deaths in the English Life Table No. 4 (males) and succeeded in breaking up the compound curve into five component curves typical of old age, middle age, youth, childhood and infancy. I want to advise my readers to study this brilliant and illuminating essay, especially on account of its beautiful form of exposition which makes the whole subject appear in a most interesting light.

Speaking of this attempt by Pearson, the American actuary, Henderson, is of the opinion that „the method has not, however, been applied to other tables and it is difficult to lay a firm foundation for it, because no analysis of the deaths into natural divisions by causes or otherwise has yet been made such that the totals in the various groups would conform to these (the Pearson) frequency curves". We shall later on come back to this statement by Henderson, which we feel is a partial truth only. On the other hand, it must be admitted that the system of Pearson's types of

skew frequency curves (by this time twelve in number) are by no means easy to handle in practical work and often require a large amount of arithmetical calculation. Moreover, there seems to be no rigorous philosophical foundation for the Pearsonian types of curves, and they can at their best only be said to be exceedingly powerful and neat instruments of graduation or interpolation.

On the other hand, I am of the opinion that the goal can be reached more easily if we, instead of the Pearsonian curve types, make use of the Laplacean-Charlier and Poisson-Charlier frequency curves, which are expressed in infinite series of the form:

$$F(x) = \varphi(x) + \beta_3 \varphi^{III}(x) + \beta_4 \varphi^{IV}(x) + \cdots, \quad (2)$$

$$\text{or } F(x) = \psi(x) + \gamma_2 \Delta^2 \psi(x) + \gamma_3 \Delta^3 \psi(x) + \cdots. \quad (3)$$

These two curve types have been treated elsewhere by Gram, Charlier, Thiele, Edgeworth Jørgensen, Guldberg and other investigators, and it is therefore not necessary to dwell further upon their analytical properties, which were discussed in Chapter I.

Returning now to the general form of our d_x curve of the mortality table which we discussed above, it is readily seen that this curve has all the properties of a compound frequency curve, that

is, a curve which is composed of several minor or subsidiary frequency curves, generally skew in appearance. As proven both by Charlier and by Jørgensen, any single valued and positive compound frequency curve vanishing at both $+\infty$ and $-\infty$ can be represented as the sum of Laplacean-Charlier and Poisson-Charlier frequency curves. We know thus a priori that the d^x curve is compounded of the two types of frequency curves. But how are we to determine the separate component curves? It is readily admitted that no a priori reason will guide us here. The purely empirical observer might therefore abandon the project right here, because to all appearances it would seem hopeless to attempt a solution by purely empirical means. The positive rationalist does not despair so easily. "Very well", he says, "if we can not make further progress by purely empirical means, we are at least permitted to try deductive reasoning and attempt to bridge the gap by means of an hypothesis". The hypothesis I shall adopt is the following:

> The frequency distribution of deaths according to age from certain groups of causes of death among the survivors in a mortality table tend to cluster around certain ages in such a manner that the frequency distribution can be represented by either a Laplacean-

Charlier or a Poisson-Charlier frequency curve.

A study of mortuary records by age and cause of death immediately supports this hypothesis. We notice, for instance, that diseases such as scarlet fever, measles, whooping cough and diphtheria often cause death among children, but rarely seem to affect older people. We know, for instance, that there is a much greater probability that a 5-year old boy will die from scarlet fever than a man at the age of 40 will die from the same disease. On the other hand, there is quite a large probability that an old man at age 85 will die from diseases of the prostate gland, while such an occurrance is almost unheard of among boys. Similarly deaths from cancer and Bright's disease are very rare in youth, but quite frequent in early old age. Tuberculosis, on the other hand, causes its greatest ravages in middle life, and has but little effect upon older ages.

6. MATHEMATICAL PROPERTIES OF THE COMPONENT FREQUENCY CURVES

Leaving, however, the question of the grouping of causes of death into a limited number of typical groups to a later discussion, we shall in the meantime see how the hypothesis can carry us over the difficulties. Let us for the moment

assume that we are able to group the causes of death into say 7 or 8 groups. We shall also assume that we know the *percentage frequency distribution* of deaths according to age in each of the groups. This means in other words that we know the equation of the frequency curves giving the percentage distribution. Let the analytical expression for these frequency curves be denoted by the symbols:

$$F_{\mathrm{I}}(x),\ F_{\mathrm{II}}(x),\ F_{\mathrm{III}}(x),\ \ldots,\ F_{\mathrm{VIII}}(x). \qquad (4)$$

Again, let the total number of deaths among the survivors in the mortality table from causes of death according to the above grouping be denoted by

$$N_{\mathrm{I}},\ N_{\mathrm{II}},\ N_{\mathrm{III}},\ N_{\mathrm{IV}},\ \ldots,\ N_{\mathrm{VIII}} \text{ respectively.} \qquad (5)$$

The number of deaths in a certain age interval, say between 50-54 can then be expressed as follows:

$$\left.\begin{aligned}\sum_{x=50}^{x=54} d_x = \sum_{50}^{54} N_{\mathrm{I}} F_{\mathrm{I}}(x) + \sum_{50}^{54} N_{\mathrm{II}} F_{\mathrm{II}}(x) + \ldots \\ + \sum_{50}^{54} N_{\mathrm{VIII}} F_{\mathrm{VIII}}(x).\end{aligned}\right\} \qquad (6)$$

In this relation the only known quantities are the equations for the frequency curves $F_{\mathrm{I}}(x)$,

$F_{II}(x)$, . . ., $F_{VIII}(x,$ of the percentage frequency distribution according to age in each of the eight groups. Neither d_x nor any of the various N's are known. The only relation we know a priori among the quantities N is the following:

$$N_I + N_{II} + N_{III} + \ldots N_{VIII} = 1{,}000{,}000. \quad (7)$$

The latter equation is simply a mathematical expression for the simple fact that the sum total of the sub-totals of the various groups of causes of death, in other words the deaths from all causes among the survivors in the mortality table, must equal the radix of the entrants of our original cohort of 1,000,000 lives at age 10. Viewed strictly from the standpoint of frequency curves, we might express the same fact by saying that the sum of the areas of the various component curves must equal 1,000,000.

It is readily seen that on the assumption that the expressions of the different $F(x)$ conform to the above hypothesis it is possible to find d_x for any age or age interval if we can determine the values of the different N's. It is in this possibility that the importance of the proposed method lies, and we shall now show how it is possible to determine the N's without knowing the exposed to risk.

7. OBSERVATION EQUATIONS Consider for the moment the following expression:

$$ {}^{54}_{50}R_{\text{III}}(x) = \frac{\sum_{50}^{54} N_{\text{III}} F_{\text{III}}(x)}{\left.\begin{aligned} &\sum_{50}^{54} N_{\text{I}} F_{\text{I}}(x) + \sum_{50}^{54} F_{\text{II}}(x) N_{\text{II}} + \\ &+ \sum_{50}^{54} N_{\text{III}} F_{\text{III}}(x) + \ldots + \sum_{50}^{54} N_{\text{VIII}} F_{\text{VIII}}(x) \end{aligned}\right\}} \qquad (8) $$

What does this equation represent? Simply the proportionate ratio of deaths in group III to the total number of deaths in all type groups (in other words the deaths from all causes) in the age interval 50-54. Such ratios are usually known as proportional death ratios. It is readily seen that these proportionate death ratios are dependent on the deaths alone and absolutely independent of the number exposed to risk, provided tne total number of deaths from all causes in a certain age group is large enough to eliminate variations due to random sampling.[1] In other words, we can find

[1] Strictly speaking this statement is only true for an age interval of one year or less and may in the case of large perturbing influences in the population exposed to risk be subject to appreciable errors when we use large age intervals of 10 or more in our grouping for the computing of $R(x)$. When the age interval for the grouping of causes of deaths by attained ages is 5 years or less the error committed in assuming $R(x)$ as being indepen-

a numerical value for the term $R_{\text{III}}(x)$ on the left side of the equation from our death records alone without reference to the exposed to risk in this interval. Similar proportionate death ratios can of course without difficulty be determined for the other groups of causes of death and for arbitrary ages or age intervals. In this manner we can determine a system of observation equations with known numerical values of $R_i(x)(i = \text{I, II, III, } \ldots)$ The fact that the number of observation equations in this system is much larger than the number of the unknown N's makes it possible to determine these unknowns by the method of least squares.

Probably the simplest manner is first to determine by simple approximation methods, or by mere inspection, approximate values for the various N's and then make final adjustments by the method of least squares.

Let, for instance,

$$'N_{\text{I}},\ 'N_{\text{II}},\ 'N_{\text{III}},\ \ldots$$

---

dent of the number exposed to risk is in most cases negligible. One of the difficulties encountered in the construction of a mortality table for Massachusetts Males was that the age interval used for the grouping was 10 years instead of 5 years or less. See in this connection the remarks at the beginning of paragraph 11 and at the conclusion of paragraph 16 of the present chapter.

be the first approximations of the areas of the various groups of frequency curves so that

$$\left.\begin{array}{c} N_{\mathrm{I}} = \alpha_1 {}'N_{\mathrm{I}},\ N_{\mathrm{II}} = \alpha_2 {}'N_{\mathrm{II}}, \ldots, \\ N_{\mathrm{VIII}} = \alpha_8 {}'N_{\mathrm{VIII}}. \end{array}\right\} \quad (9)$$

Let us furthermore introduce the following symbols:

$$\left.\begin{array}{c} N_{\mathrm{I}} F_{\mathrm{I}}(x) = \Phi_1(x),\ {}'N_{\mathrm{II}} F_{\mathrm{II}}(x) = \Phi_2(x), \ldots, \\ {}'N_{\mathrm{VIII}} F_{\mathrm{VIII}}(x) = \Phi_8(x). \end{array}\right\} (10)$$

The different values of

$$\Phi_1(x),\ \Phi_2(x),\ \Phi_3(x),\ \ldots,\ \Phi_8(x)$$

may then be regarded as a system of component frequency curves to which we now must apply the different correction factors $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_8$ in order to fit the curves to the observed proportional death ratios, $R(x)$, for the various groups of typical causes of death. Let us for example assume that the observed death ratio of a certain age (or age group), x, under a certain group of causes of death, say group No. III, is $R_{\mathrm{III}}(x)$. We have then the following observation equation:

$$\left.\begin{array}{c} R_{\mathrm{III}}(x) = \alpha_3 \Phi_3(x) : [\alpha_1 \Phi_1(x) + \alpha_3 \Phi_3(x) + \\ + \alpha_4 \Phi_4(x) + \ldots + \alpha_8 \Phi_8(x) + \alpha_2 \Phi_2(x)] \end{array}\right\} (11)$$

Since the sum of the areas of the different component curves necessarily must equal 1,000,000 it is easy to see that we may write the factor α_2 in the last term of the denominator in the following form:

$$\alpha_2 \sum' \Phi_2(x) = 1{,}000{,}000 - \left[\alpha_1 \sum' \Phi_1(x) + \alpha_3 \sum' \Phi_3(x) + \ldots + \alpha_8 \sum' \Phi_8(x)\right]$$

or

$$\alpha_2 = \left(1{,}000{,}000 - \left[\alpha_1 \sum' \Phi_1(x) + \alpha_3 \sum' \Phi_3(x) + \ldots + \alpha_8 \sum' \Phi_8(x)\right]\right) : \sum' \Phi_2(x) = k_0 - \left[k_1\alpha_1 + k_3\alpha_3 + \ldots + k_8\alpha_8\right]$$

where

$$\left.\begin{aligned} k_0 &= \frac{1{,}000{,}000}{\Sigma\,\Phi_2(x)}, \quad k_1 = \frac{\Sigma\,\Phi_1(x)}{\Sigma\,\Phi_2(x)}, \\ k_1 &= \frac{\Sigma\,\Phi_3(x)}{\Sigma\,\Phi_2(x)}, \ \ldots, \ k_8 = \frac{\Sigma\,\Phi_8(x)}{\Sigma\,\Phi_2(x)}. \end{aligned}\right\} \quad (12)$$

The expression for $R_{III}(x)$ can then be put in the following form:

$$\left.\begin{aligned} R_{III}(x) = \alpha_3\Phi_3(x) : [&\alpha_1\Phi_1(x) + \alpha_3\Phi_3(x) + \\ &+ \alpha_4\Phi_4(x) + \ldots + \alpha_8\Phi_8(x) + \\ &+ (k_0 - k_1\alpha_1 - \ldots - k_8\alpha_8)\,\Phi_2(x)]. \end{aligned}\right\} \quad (13)$$

Similar observation equations for the other groups are derived without difficulty.

Once having formed the observation equations it is simply a matter of routine work to compute the normal equations from which the values of the unknown N's can be found. We shall, however, not go into detail with the derivation of the necessary formulas, since this is a process which belongs wholly to the domain of the theory of least squares and which has received adequate treatment elsewhere. (See for instance Brunt's *Combination of Observations.*)

8. CLASSIFICATION OF CAUSES OF DEATH

We think it more advantageous to illustrate the method by a concrete example. As an illustration we may take the case of Michigan Males in the period 1909—1915. The mortuary records of Males in Michigan are for that period given in the reports issued annually by the Secretary of State on "Registration of Births and Deaths, Marriages and Divorces in Michigan". The deaths by sex, age and cause of death are given in quinquennial age groups. A very serious drawback is the grouping of all ages above 80 into a single age group instead of in at least 4 or 5 quinquennial age groups. This makes it impossible to obtain good observation equations

for ages above 80. When we consider that about one fifth of the original entrants at age 10 in the mortality table die after the age of 80, it is readily seen that this defect in the Michigan data is of a very serious character, which makes it out of the question to determine correctly the areas of the curves for middle old age and extreme old age. For ages below 70 these curves do not play so important a role, and the method ought therefore in these ages yield satisfactory results. We now make the assertion that the deaths among the survivors in the final life table can be grouped in the following typical groups.

Causes of Death typical of :—	
Group I	Extreme Old Age.
— II	Middle Old Age.
— III	Early Old Age.
— IV	Middle Life.
— V	Early Middle Life.
— VI	Pulmonary Tuberculosis, Etc.
— VIIa	Early Life Occupational Hazard.
— VIIb	Middle Life Occupational Hazard.
— VIIIa	Childhood.

The classification of causes of death according to this scheme is given in the following table, marked Table A.

Table A. Michigan Males 1909—1915
Classification of causes of death according to the chosen system of curves.

No. in International Classification.	GROUP I
81.	Diseases of the arteries.
124.	Diseases of the bladder.
125—133.	Other diseases of the genito-urinary system.
142.	Gangrene.
154.	Old age.
126.	Diseases of the prostate.

GROUP II

10.	Influenza.
47—48.	Rheumatism.
64.	Apoplexy.
65.	Softening of the brain.
66.	Paralysis.
79.	Heart disease.
82.	Embolism.
89.	Acute bronchitis.
90.	Chronic bronchitis.
91.	Broncho-pneumonia.
94.	Congestion of the lungs.
96—97.	Asthma and emphysema.
103.	Other diseases of the stomach.

No. in International Classification.

105. Diarrhea and enteritis. (over 2 years)
14. Dysentery.

GROUP III

39. Cancer of the mouth.
40. Cancer of the stomach and liver.
41. Cancer of the intestines.
44. Cancer of the skin.
45. Cancer af other organs.
46. Tumors.
50. Diabetes.
53—54. Leukemia and anemia.
63. Other diseases of the spinal cord.
68. Other forms of mental diseases.
80. Angina pectoris.
109—110. Hernia, intestinal obstruction, and other diseases of the intestines.
120. Bright's disease.
121. Other diseases of the kidneys
123. Calculi of urinary passages.

GROUP IV

56. Alcoholism.
18. Erysipelas.
62. Locomotor ataxia.
73—76. Other diseases of the nervous system.
77. Pericarditis.

No. in International Classification

78. Endocarditis.
83. Diseases of the veins.
84. Diseases of the lymphatics.
85—86. Other diseases of the circulatory system.
87. Diseases of the larynx.
88. Diseases of the thyroid body.
92. Pneumonia.
93. Pleurisy.
95. Gangrene of the lungs.
98. Other diseases of the respiratory system.
99—101. Diseases of the mouth, pharynx, and oesophagus.
111. Acute yellow atrophy of the liver.
113. Cirrhosis of the liver.
114. Biliary calculi.
115—116. Diseases of the liver and spleen.
118. Other diseases of the digestive system.
143—145. Furuncle, abscess, and other diseases of the skin.
147—149. Diseases of the joints, and locomotor system.

GROUP V

4. Malarial fever.
13. Cholera nostras.

No. in International Classification.	
20.	Septicemia.
24.	Tetanus.
32.	Pott's disease.
33.	White swellings.
34.	Tuberculosis of other organs.
35.	Disseminated tuberculosis.
55.	Other general diseases.
60.	Encephalitis.
70—71.	Convulsions.
102.	Ulcer of the stomach.
117.	Peritonitis.
119.	Acute Nephritis.
164.	Diseases of the bones.
155.	Suicide by poison.
156.	Suicide by asphyxia.
157.	Suicide by hanging.
158.	Suicide by drowning.
159.	Suicide by firearms.
160.	Suicide by cutting instruments.
161.	Suicide by jumping from hight places.
163.	Suicide by other or unspecified means.
164—165.	Accidental poisonings.
166.	Conflagration.
167.	Burns (conflagration excepted).
168.	Inhalation of noxious gases.
172.	Traumatism by fall.

No. in International Classification.

175—(2). Traumatism by electric railway.
175—(3). Traumatism by automobiles.
175—(4). Traumatism by other vehicles.
176. Traumatism by animals.
178. Cold and freezing.
179. Effects of heat.
185. Fractures and dislocations (cause not specified.

GROUP VI

28. Tuberculosis of the lungs.
29. Miliary tuberculosis.
37—38. Venereal diseases.
186. Other accidental traumatism.
57—59. Chronic poisoning.
67. General paralysis of the insane.
31. Abdominal tuberculosis.

GROUP VII

1. Typhoid fever.
69. Epilepsy.
108. Appendicitis.
182. Homicide.
169. Accidental drowning.
170. Traumatism by firearms.
171. Traumatism by cutting instruments.

No. in International Classification.	
173.	Traumatism by mines and quarries.
174.	Traumatism by machinery.
175—(1).	Traumatism by railroads.
180.	Ligthning.
61.	Meningitis.

GROUP VIII

5.	Smallpox.
6.	Measles.
7.	Scarlet fever.
8.	Whooping cough.
9.	Diphtheria and croup.
30.	Tubercular meningitis.
150.	Congenital malformations.

9. OUTLINE OF COMPUTING SCHEME The number of deaths in the various groups according to the above classification and arranged according to age during the period 1909—1915 is given in the table B on page 140.

From that table it is a simple matter to compute the proportionate death ratios of the separate groups of causes of death. Such a computation is shown in table C on page 141.

It is readily seen that these death ratios are independent of the number exposed to risk. More-

over, the number of observations seem to be sufficiently large to eliminate serious variations due to random sampling. This might perhaps not hold true for the age intervals 10 to 14 and 15 to 19 where not alone random sampling is present, but a somewhat modified classification seems necessary. I have, however, not used the observed proportionate death ratios for the two younger age intervals in my computations which only took into account the ratios above 20. For this reason I do not deem it necessary to go into a closer investigation of a re-classification of causes of death for these younger age groups. A more serious defect which cannot be overcome is presented in the ages above 80 where, as mentioned before, a classification according to age is absent in the original records for the state of Michigan. The fact that the highest number of deaths (12,473) occurred in ages above 80 makes this defect more serious than the omission of a re-classification of causes of death below 20.

So far we have only been concerned with the first step in the complete induction according to the model of Jevons, namely that of simple observation. The next step in the induction is the hypothesis. We present now the following working hypothesis.

The frequency distribution of deaths according

Table B. Michigan 1909—15 (Males)

Number of Deaths according to Groups of Causes of Death.

	I	II	III	IV	V	VI	VII	VIII	Total
10—14			186	198	359	129	683	558	2113
15—19			196	343	406	712	1106	411	3174
20—24			299	452	664	1178	1638	426	4657
25—29		423	331	547	710	1223	1489	76	4799
30—34		542	407	646	668	1197	1005	56	4521
35—39	29	708	515	784	712	1143	793	41	4725
40—44	47	894	722	847	690	906	637	40	4783
45—49	78	1352	1106	1023	715	861	553	22	5710
50—54	127	1890	1595	1123	808	786	468	16	6813
55—59	182	2471	1940	1122	750	630	341	12	7448
60—64	353	3450	2346	1110	651	477	282	12	8681
65—69	740	4590	2592	1096	596	356	208	6	10184
70—74	1275	4996	2418	1000	510	253	142	2	10596
75—79	1811	4655	1946	799	391	149	92	5	9848
80 & over.....	4165	5192	1706	749	473	85	101	2	12473
Total	8807	31163	18305	11839	9103	10085	9538	1685	100525

Tabel C. Proportionate Death Ratios or R_x

	I	II	III	IV	V	VI	VII	VIII	Total
10—14			8.8	9.4	17.0	6.1	32.3	26.4	100.0
15—19			6.2	10.8	12.8	22.4	34.8	13.0	100.0
20—24		0.6	5.8	9.7	14.3	25.2	35.2	9.2	100.0
25—29		8.5	6.6	11.4	14.8	25.5	31.0	2.2	100.0
30—34		12.0	9.0	14.3	14.8	26.5	22.2	1.2	100.0
35—39	0.6	15.0	10.9	16.6	15.5	24.2	16.8	0.8	100.0
40—44	1.0	18.7	15.1	17.7	14.4	19.0	13.6	0.5	100.0
45—49	1.4	23.7	19.3	17.9	12.5	15.1	9.7	0.4	100.0
50—54	1.9	27.8	23.4	16.5	11.9	11.5	7.0	0.0	100.0
55—59	2.4	33.2	26.0	15.1	10.1	8.4	4.6	0.2	100.0
60—64	4.1	39.7	27.0	12.8	7.5	5.5	3.3	0.1	100.0
65—69	7.3	45.1	25.5	10.8	5.8	3.5	2.0	0.0	100.0
70—74	12.0	47.2	22.8	9.4	4.7	2.4	1.4	0.0	100.0
75—79	18.4	47.3	19.8	8 1	4.0	1.5	0.9	0.0	100.0
80 & over.....	33.4	41.6	13.7	6.0	3.8	0.7	0.8	0.0	100.0

to age of the above groups of causes of death among the survivors of an original cohort of 1,000,000 entrants at age 10 can be represented by a system of frequency curves determined by the following characteristic parameters:

Parameters

Group	Mean	Dispersion	Skewness	Excess
I	79.5 years	9.5730 years	+.1056	+.0546
II	70.5 —	12.8000 —	+.0967	+.0126
III	65.5 —	13.6870 —	+.1248	+.0650
IV	59.5 —	17.0890 —	+.1790	−.0106
V	55.5 —	19.9411 —	+.0555	−.0367
VI	44.5 —	16.0352 —	−.0124	−.0272
VIIb	57.5 —	12.1552 —	+.0008	−.0005
VIIa	Poisson-Charlier Curve: Modulus = 28.5 years, Eccentricity = 1.0001			
VIIIa	Poisson-Charlier Curve: Modulus = 13.5 years.			

From these parameters and from well-known tables of the probability or normal frequency curve and its various derivatives it is easy to determine the frequency distribution for any desired interval.

For this system of frequency curves we now shall try to find the various areas of N_{I}, N_{II}, N_{III}, ... , N_{VIII} so as to conform to the observed values of R_x in Table C. As a first approach to the final values of N, we may by an inspection (which of course is improved upon by

a long practice in curve fitting) choose the following approximations.[1]

Group	Approximate Value of 'N.
I	123000
II	366000
III	183000
IV	105000
V	75000
VI	70000
VIIa & VIIb	61000
VIII	17000
	1000000

These preliminary numerical values represent the first approximations of the areas of the various frequency curves. The sequence represented by

$$'N_{\mathrm{I}}F_{\mathrm{I}}(x), 'N_{\mathrm{II}}F_{\mathrm{II}}(x), 'N_{\mathrm{III}}F_{\mathrm{III}}(x), \cdots 'N_{\mathrm{VIII}}F_{\mathrm{VIII}}(x) \quad (14)$$

gives the number of deaths at age x. We notice thus that by multiplying the various equations of frequency curves for arbitrary age intervals with

[1] These numbers represent as a matter of fact a first rough approximation of the areas of the different component curves by means of the method of point contours. Hence it is to be expected that the final adjustments will be comparatively small. This fact has, however, no influence upon the application of the method.

Table D.

Michigan Life Table for Males 1909—1915. (1st Approximation).

Ages	I	II	III	IV	V	VI	VIIa	VIIb	VIII	dx
10—14.....			212	817	931	1247	1656	2	8312	13191
15—19.....		29	553	1316	1506	2362	5012	14	5879	16671
20—24.....		306	1159	1910	2219	3783	8068	48	2170	19663
25—29.....		836	1993	2532	3025	5304	8879	150	530	23251
30—34.....		1973	2793	3153	3847	6649	7412	386	96	26307
35—39.....		3894	3270	3857	4623	7604	4972	826	13	29059
40—44.....	177	6692	3712	4825	5335	8032	2787	1491	2	33050
45—49.....	646	10544	5510	6333	5943	7933	1342	2271		40522
50—54.....	1576	16339	10453	8415	6452	7355	564	2941		54096
55—59.....	2664	25537	18792	10727	6801	6354	211	3236		74322
60—64.....	3672	38552	27641	12575	6903	5078	72	3024		97517
65—69.....	6645	52153	32382	13216	6669	3700	22	2399		117186
70—74.....	14853	60162	30173	12266	6050	2355	3	1616		127478
75—79.....	25988	57361	22338	9929	5100	1317		914		122947
80—84.....	29856	44566	12936	6890	3922	622		436		99228
85—89.....	22200	27772	5838	3966	2728	232		173		62909
90—94.....	10559	13497	1991	1733	1682	57		54		29573
95—99.....	3173	4800	593	377	881	2		14		9844
100 & over	1081	982	706		413			4		3186
	123089	365995	183045	104838	75030	69986	41000	20003	17002	1000000
	$\Sigma\Phi_1$	$\Sigma\Phi_2$	$\Sigma\Phi_3$	$\Sigma\Phi_4$	$\Sigma\Phi_5$	$\Sigma\Phi_6$	$\Sigma\Phi_7$		$\Sigma\Phi_8$	

their respective 'N's we can get a first approximation of the final death curve. I give on page 144 an approximate table arranged in 5 year intervals.

We might now first compute the various factors k_0, k_3 ..... k_8 which will be common for all observation equations. We have, referring to the above formulas (11 and 12) for the various k's (15).

$$\left.\begin{aligned} k_0 &= \frac{1000000}{365995}; \quad k_1 = \frac{123089}{365995}; \quad k_3 = \frac{183045}{365995}; \\ k_4 &= \frac{104888}{365995}; \quad k_5 = \frac{75030}{365995}; \quad k_6 = \frac{69996}{365995}; \\ k_7 &= \frac{61003}{365995}; \quad k_8 = \frac{17002}{365995}. \end{aligned}\right\} \quad (15)$$

Or

$$\left.\begin{aligned} &k_0 = 2{,}732, \; k_1 = 0{,}336, \; k_3 = 0{,}500, \; k_4 = 0{,}287, \\ &k_5 = 0{,}205, \; k_6 = 0{,}191, \; k_7 = 0{,}167, \; k_8 = 0{,}046. \end{aligned}\right\}$$

To illustrate the further process of the computation of the observation equations, let us take a certain age interval, say the interval between 50-54. The value of Φ_2 taken from the above table is 163.39. The value of $R_{III}(x)$ for this interval is 0.234 (see table page 141). Hence we have the following observation equation (16).

$$\left.\begin{array}{l} 0.234 = 104.53\alpha_3 : [15.76\alpha_1 + 104.53\alpha_3 + \\ 84.16\alpha_4 + 64.52\alpha_5 + 73.55\alpha_6 + 35.01\alpha_7 + \\ 0.00\alpha_8 + (2.732 - 0.336\alpha_1 - 0.500\alpha_3 - \\ 0.287\alpha_4 - 0.205\alpha_5 - 0.191\alpha_6 - 0,167\alpha_8 - \\ \qquad - 0.046\alpha_8)\ 163.39]. \end{array}\right\} (16)$$

After a few simple reductions this may be brought to the following form:

$$\left.\begin{array}{l} 9.16\alpha_1 + 99.19\alpha_3 - 8.72\alpha_4 - 7.26\alpha_5 - \\ 9.91\alpha_6 - 1\text{-}81\alpha_7 + 1.76\alpha_8 - 104.45 = 0. \end{array}\right\} (17)$$

In the routine work I usually use a system of computing the various equations which is outlined in detail in the accompanying tabular scheme referring to all the groups in the age interval 50-54 and shown on pages 148-154.

Similar observation equations are arrived at in exactly the same manner for other groups and other age intervals. For the whole interval from age 20 and upwards we get in this way 96 observation equations from which to determine the correction factors. The coefficients of these observational equations are then written down, and

their various products formed in turn. We deem it not necessary to give all these observational equations and their coefficients for all the 96 observations, but shall limit ourselves to give all the necessary computations for the interval from 50-54 as previously considered. With the usual system of notation employed in the method of least squares we get the scheme on pages 148-154.

Normal Equations, Michigan Males 1909—1915.

723763	400750	218930	150776	135184	115318	30325	1801152
	877847	253187	176242	149858	129697	34600	2053941
		237159	90440	72317	62110	16246	964843
			105346	47022	39939	10576	628608
				76774	28909	8668	525295
					53378	7012	437390
						2391	111625

The addition of the various columns of the sum products of the coefficients gives us finally the above set of normal equations of which we only submit the coefficients in the usual scheme employed in the method of least squares.

Solving the above system of normal equations by means of the well-known method devised by Gauss, we obtain finally the values on page 154 for the various α's by which the approximate values $'N$ must be multiplied in order to yield the probable values of N.

Scheme for Calculation of the Observational Equations. Michigan, Males, 1909-1915. Age Interval (50-54)

Group	r	(1) $\Phi_2(x)$	(2) (1)×(9)	(3) $\Phi_r(x)$	(4) (2)+(3)	(5) $\mathbf{R}(x)$	(6) (4)×(5)	(7) $\Phi_r(x)$	(8) (7)−(6)	(9) Common Factor
I	0	163.39	+446.38		+446.38		+ 8.48		− 8.48	+ 2.732
	1		− 54.90	15.76	− 39.14	.019	− .74	15.76	+ 16.50	− 0.336
	3		− 81.70	104.53	+ 22.83		+ .43		− .43	− 0.500
	4		− 46.89	84.16	+ 37.27		+ .71		− .71	− 0.287
	5		− 33.49	64.52	+ 31.03		+ .59		− .59	− 0.205
	6		− 31.21	73.55	+ 42.34		+ .80		− .80	− 0.191
	7		− 27.29	35.01	+ 7.72		+ .15		− .15	− 0.167
	8		− 7.52	0.00	− 7.52		− .14		+ .14	− 0.046
III	0						+104.45		−104.45	
	1						− 9.16		+ 9.16	
	3					.234	+ 5.34	104.53	+ 99.19	
	4						+ 8.72		− 8.72	
	5						+ 7.26		− 7.26	
	6						+ 9.91		− 9.91	
	7						+ 1.81		− 1.81	
	8						− 1.76		+ 1.76	
IV	0						+ 73.65		− 73.65	
	1						− 6.46		+ 6.46	
	3						+ 3.77		− 3.77	
	4					.165	+ 6.15	84.16	+ 78.01	
	5						+ 5.12		− 5.12	
	6						+ 6.99		− 6.99	
	7						− 1.27		− 1.27	
	8						− 1.24		+ 1.24	

V	0						−	53.11	
	1						+	4.66	
	3						−	2.72	
	4						−	4.44	
	5		.119		64.52		+	60.83	
	6						−	5.04	
	7						−	0.92	
	8						+	0.89	
		+	53.11						
		−	4.66						
		+	2.72						
		+	4.44						
		+	3.69						
		+	5.04						
		+	0.92						
		−	0.89						
VI	0						+	51.33	
	1						−	4.50	
	3						+	2.63	
	4						−	4.29	
	5		.115		73.55		−	3.57	
	6						+	68.68	
	7						−	0.89	
	8						+	0.86	
		+	51.33						
		−	4.50						
		+	2.63						
		+	4.29						
		+	3.57						
		+	4.87						
		−	0.89						
		−	0.86						
VII	0						+	31.25	
	1						−	2.74	
	3						−	1.60	
	4						−	2.61	
	5						−	2.17	
	6		.070		35.01		−	2.96	
	7						+	34.47	
	8						+	0.53	
		+	31.25						
		−	2.74						
		+	1.60						
		+	2.61						
		+	2.17						
		+	2.96						
		+	0.54						
		−	0.53						
II	0								(2)−(6)
	1		.278						+322.29
	3								44.02
	4								88.05
	5								57.25
	6								42.12
	7								42.98
	8								29.44
		+	124.09						6.43
		−	10.88						
		+	6.35						
		+	10.36						
		+	8.63						
		+	11.77						
		+	2.15						
		−	2.09						

Note: — In the tables on pages 151—154 the plus and minus signs of the various products have been omitted.

	a	b	c	d	e	f	g	h	s
20—24	16.5	−0.4	−0.7	−0.6	−0.8	−0.2	0.1	−8.5	5.4
	9.2	99.2	−8.7	−7.3	−9.9	−1.8	1.8	−104.5	−22.0
	6.5	−3.8	78.0	−5.1	−7.0	−1.3	1.2	−73.6	−5.1
50—54	4.7	−2.7	−4.4	60.8	−5.0	−0.9	0.9	−53.1	0.3
	4.5	−2.6	−4.3	−3.6	68.7	−0.9	0.9	−51.3	11.4
	2.7	−1.6	−2.6	−2.2	−3.0	34.5	0.5	−31.3	−3.0
80 & over	−44.0	−88.1	−57.3	−42.1	−43.1	−29.4	−5.4	322.3	12.9

	aa	ab	ac	ad	ae	af	ag	ah	as
	272.3	6.6	11.6	9.9	13.2	3.3	1.7	140.3	89.1
	84.6	912.6	80.0	67.2	91.1	16.6	16.6	961.4	202.4
	42.3	24.7	507.0	33.2	45.5	8.5	7.8	478.4	33.2
50—54	22.1	12.7	20.7	285.8	23.5	4.2	4.2	249.6	1.4
	20.4	11.7	19.4	16.2	309.2	4.1	4.1	230.9	51.3
	7.3	4.3	7.0	5.9	8.1	93.2	1.4	84.5	8.1
	1936.0	3876.4	2421.2	1852.4	1896.4	1293.6	237.6	14181.2	567.6
Sum:	723763.0	400750.0	218930.0	150776.0	135184.0	115318.0	30325.0	−1801152.0	−26106.0

	bb	bc	bd	be	bf	bg	bh	bs
	0.2	0.3	0.2	0.3	0.1	0.0	3.3	2.2
	9840.6	863.0	724.2	982.1	178.6	178.6	10366.4	2182.4
	14.4	296.4	19.4	26.6	4.9	4.6	279.7	19.4
50—54	7.3	11.9	164.2	13.5	2.4	2.4	143.4	0.8
	6.8	11.2	9.4	178.6	2.3	2.3	133.4	29.6
	2.6	4.2	3.5	4.8	55.2	0.8	50.1	4.2
	7761.6	5048.1	3709.0	3797.1	2590.1	474.7	28394.6	1136.5
Sum:	877847.0	253187.0	176242.0	149858.0	129697.0	34600.0	−2053941.0	−31760.0

	cc	cd	ce	cf	cg	ch	cs
	0.5	0.4	0.6	0.1	0.1	6.0	3.8
	75.7	63.5	86.1	15.7	15.7	909.2	191.4
	6084.0	397.8	546.0	101.4	93.6	5740.8	397.8
50—54	19.4	267.5	22.0	4.0	4.0	233.6	1.3
	18.5	15.5	295.4	3.9	3.9	220.6	49.0
	6.8	5.7	7.8	89.7	1.3	81.4	7.8
	3283.3	2412.3	2469.6	1684.6	309.4	18467.8	739.2
Sum:	237159.0	90440.0	72317.0	61110.0	16246.0	−964843.0	−14454.0

	dd	de	df	dg	dh	ds
	0.4	0.5	0.1	0.1	5.1	3.2
	53.3	72.3	13.1	13.1	762.8	160.6
	26.0	35.7	6.6	6.1	375.4	26.0
50—54	3696.6	304.0	54.7	54.7	3228.5	18.2
	13.0	247.3	3.2	3.2	184.7	41.0
	4.8	6.6	75.9	1.1	68.9	6.6
	1772.4	1814.5	1237.7	227.3	13568.8	543.1
Sum:	105436.0	47022.0	39939.0	10576.0	−628608.0	−8177.0

	ee	ef	eg	eh	es
	0.6	0.2	0.1	6.8	4.3
	98.0	17.8	17.8	1034.6	217.8
	49.0	9.1	8.4	515.2	35.7
50—54	25.0	4.5	4.5	265.5	1.5
	4719.7	61.8	61.8	3524.3	783.2
	9.0	103.5	1.5	93.9	9.0
	1857.6	1267.1	232.7	13891.1	556.0
Sum:	76774.0	28908.0	8669.0	−525293.0	−6562.0

	ff	fg	fh	fs
	0.0	0.0	1.7	1.1
	3.2	3.2	188.1	39.6
	1.7	1.6	95.7	6.6
50—54	0.8	0.8	47.8	0.3
	0.8	0.8	46.2	10.3
	1190.3	17.3	1079.9	103.5
	864.4	158.8	9475.6	379.3
Sum:	53378.0	7012.0	−437389.0	−1027.0

	gg	gh	gs
	0.0	0.8	0.5
	3.2	188.1	39.6
	1.4	88.3	6.1
50—54	0.8	47.8	0.3
	0.8	46.2	10.3
	0.3	15.7	1.5
	29.2	1740.4	69.7
Sum:	2391.0	— 111625.0	—1807.0

	hh	hs
	72.3	45.9
	10920.3	2299.0
	5416.9	375.4
50—54	2819.6	15.9
	2631.7	584.8
	979.7	93.9
	103877.3	4157.7
Sum:	6630212.0	107358.0

Correction Factors. α.

Group	I	1.03284
—	II	1.00017
—	III	1.03635
—	IV	1.03731
—	V	1.00956
—	VI	0.97334
—	VIIa	0.90332
—	VIIb	0.60565
—	VIII	1.13743

Applying the above correction factors to the respective values of $'N$, we get finally as the total areas of the respective component curves:

Group	I	127,131
—	II	366,059
—	III	189,699
—	IV	108,750
—	V	75,747
—	VI	68,130
—	VIIa	33,032
—	VIIb	12,133
—	VIII	19,339
		1,000,000

Multiplying the equations of the various frequency curves, $F(x)$, of the percentage distribution in each group with the above values of N we obtain finally the complete mortality table as will be given in the Appendix. The final graphical representation of the frequency curves is shown in Figure 2.

10. GOODNESS OF FIT

This completes the third step in the inductive process. The fourth and final step is the verification of the results thus arrived at by a mere deductive process. Here it must be remembered

Fig. 2. Michigan Life Table.

that the condition which the final component frequency curves shall fulfill is the one that observed proportionate death ratios shall agree as closely as possible with the expected or theoretical proportionate death ratios as computed from the final table. In this connection it must be borne in mind that the observed proportionate death ratios are given in quinquennial age groups. Thus the observed proportionate death ratios in a certain age interval, as for example between 50—54 are really the average or "central" proportionate death ratios at age 52. From the complete table it is, however, possible to compute the proportionate death ratios for each specific age. Graphically the expected proportionate death ratios will therefore represent a continuous curve, while the observed ratios will be represented by a rectangular shaped column diagram. Such a graphical representation is shown in Fig. 3 which simply represents the figures in Table C and Table E in graphical form. The "goodness of fit" of the "expected" or theoretical values to the "actual" or observed values is seen to be very close, especially in the largest and most important groups. It is only in the combined groups VIIa and VIIb that the "fit" might probably be open to criticism for higher ages, but even here the deviation is small between the actual and theoretical values. A very small increase in the

Table E.

Expected or theoretical Proportionate Death Ratios as computed from the proposed system of frequency curves for Michigan Males 1909—1915.

Age	I	II	III	IV	V	VI	VII	VIII
10			1.0	5.6	6.2	7.3	3.9	75.9
15			2.4	7.0	7.8	11.2	17.7	53.8
20		1.1	5.2	9.9	11.1	17.8	33.0	21.9
25		2.7	8.4	11.7	13.4	22.6	36.1	5.1
30		5.9	11.0	12.8	15.1	25.4	28.9	0.9
35		11.2	12.2	13.8	16.4	26.7	19.5	0.1
40	0.2	18.1	11.9	15.1	17.0	25.7	11.9	0.0
45	1.1	24.5	12.8	16.3	16.1	21.9	7.3	
50	2.5	29.0	17.4	16.6	13.5	15.9	3.9	
55	3.6	32.6	24.1	15.7	10.4	10.2	3.4	
60	3.8	36.6	28.6	14.1	7.9	6.2	2.3	
65	4.4	42.4	29.4	12.3	6.2	3.8	1.5	
70	8.7	46.2	26.3	10.6	5.1	2.2	0.9	
75	17.5	46.6	20.9	8.8	4.3	1.3	0.5	
80	27.7	45.0	14.9	7.5	3.9	0.8	0.3	
85	34.7	43.4	10.6	6.6	4.1	0.4	0.2	

area of the VIIb curve would easily adjust this difference. It is, however, doubtful if such a correction or adjustment would have any noteworthy effect upon the ultimate mortality rates q_x, and I do not consider it worth while to go to the additional trouble of recomputing the areas, especially in view of the fact that the observation data above the age of 80 are not exact and detailed enough to be used in this method of curve fitting. For ages up to 70 or 75 I consider, however, the table as thus constructed as sufficiently accurate for all practical purposes.

11. MASSACHUSETTS MALES 1914—1917 As another example of the method I take the construction of a mortality table for the State of Massachusetts from the mortuary records for the three years 1914, 1915 and 1916. The records as given by the Registration reports are better than the records for Michigan, in as much as they have avoided the deplorable practice of grouping all deaths above the age of 80 into a single age group. On the other hand, the classifications of cause of death in Massachusetts by attained age are given in ten year age groups only. Hence it is readily seen that we will only be able to secure half as many observation equations as in the case of the five year interval in Michigan.

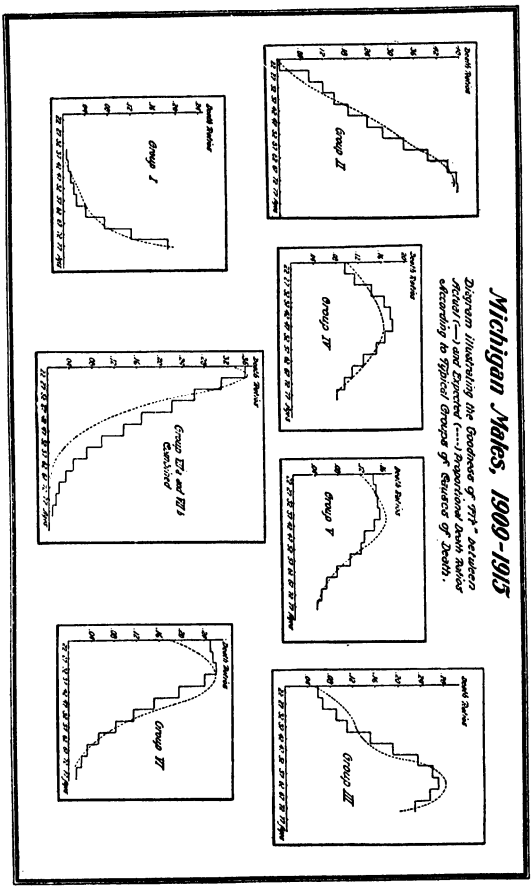

Fig. 3

This rather large grouping puts the method to a severe test. In spite of this drawback I shall for

the benefit of the readers briefly outline the results I have obtained from an analysis of the Massachusetts data.

While for the Michigan data I employed a system of frequency curves previously used with success for certain Scandinavian data, I found it was easier to fit the Massachusetts data to a system of frequency curves used in the construction of a mortality table for England and Wales for the years 1911 and 1912 from the mortuary records of deaths by age and cause among male lives. *The classification by age of the causes of death in 8 groups is also different from that of Michigan, especially for middle life and younger ages.* The parameters of the system of component frequency curves to which I fitted the Massachusetts data are shown in the following table *F:*

Table F.

Parameters of the System of Frequency Curves for Massachusetts Males 1914—1916.

Group	Mean	Dispersion	Skewness	Excess
I	78.70 years	7,9775 years	+ .0920	+ .0331
II	68.00 —	12,2051 —	+ .1151	+ .0234
III	63.05 —	13,0532 —	+ .1210	+ .0471
IV	60.45 —	17,8552 —	+ .0983	— .0091
V	49.60 —	18,5100 —	+ .0328	— .0309
VI	43.80 —	14,6750 —	— .0091	— .0272
VIIb	57.40 —	12,1550 —	+ .0021	— .0025

VIIa and VIIIa constructed from Poisson-Charlier Curves.

The observed number of deaths according to the 8 groups of causes of death, and their corresponding proportionate death ratios are given in the following tables *G* and *H*.

By finding first approximate values and then by a further correction of these approximation areas by means of the factors α. determined by the method of least squares in exactly the same manner as demonstrated in the case of Michigan, we finally arrive at the following areas of the various groups.

Areas of the component frequency curves in the Life Table for Massachusetts Males, 1914—1916.

	Areas	
Group	I	90064
—	II	281470
—	III	207854
—	IV	151316
—	V	99543
—	VI	107718
VIIa &	VIIb	40719
—	VIIIa	21316
		1000000

Forming the products $N\,F(x)$ for the various groups and integral ages we obtain finally the life table as shown in the appendix. In order

Table G.

Massachusetts Males) 1914—1916.

Number of Deaths according to Groups of Causes of Death.

Ages	I	II	III	IV	V	VI	VII	VIII	Total
10—14			66	62	330	47	152	345	1002
15—19			131	112	341	384	248	320	1536
20—29		384	391	524	1025	1909	666	176	5075
30—39		838	718	992	1107	2187	528	125	6495
40—49	103	1526	1454	1449	1209	2024	492	87	8344
50—59	357	2830	2621	1788	1131	1336	331	51	10445
60—69	1073	3961	3172	1792	835	626	177	30	11666
70—79	1809	4121	2564	1432	477	182	79	21	10685
80—89	1302	1909	871	577	123	35	20	5	4842
90 & over	226	270	78	76	19	1	1		671

Table H.

Proportionate Death Ratios of Above Numbers. (R_x).

10—14			6.6	6.2	13.0	4.6	12.2	57.4	100.0
15—19			8.5	7.3	22.2	25.0	16.2	20.8	100.0
20—29		7.6	7.7	10.3	20.2	37.6	13.1	3.5	100.0
30—39		12.9	11.1	15.3	17.0	33.7	8.1	1.9	100.0
40—49	1.2	18.3	17.4	17.4	14.5	24.3	5.9	1.0	100.0
50—59	3.4	27.1	25.1	17.1	10.8	12.8	2.3	0.5	100.0
60—69	9.2	34.0	27.2	15.4	7.2	5.4	1.5	0.1	100.0
70—79	16.9	38.6	24.0	13.4	4.5	1.7	0.8	0.1	100.0
80—89	26.9	39.4	18.0	11.9	2.5	0.8	0.4	0.1	100.0
over 90	33.7	40.2	11.6	11.3	2.8	0.2	0.2	0.0	100.0

Massachusetts Males, 1914-1916
Number of Deaths at Various Ages out of an Original Group of 100,000 Entering at Age 10
Number of Deaths
2500
2000
1500
1000
500
0
Key
A – All Causes of Death
B – Causes Typical of: —
I – Extreme Old Age
II – Middle Old Age
III – Early Old Age
IV – Middle Life
V – Early Middle Life
VI – Pulmonary Tuberculosis, etc.
VII – Early Life Occupational Hazard
VIII – Middle Life Occupational Hazard
IX – Childhood
A Composite Frequency Curve
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105
Ages at Death
Constructed from Physicians' Reports of Massachusetts for the Period 1914-1916.

Fig. 4.

to test the "goodness of fit" of the curves it is necessary to compute the expected or theoretical proportional death ratios from this latter table and compare such ratios with the observed or actual proportionate death ratios as shown in Table H. The theoretical values are shown in Table I, and a graphical representation illustrating the "goodness of fit" between the observed and theoretical ratios is given in Fig. 5. I think it will be generally admitted that the fit is satisfactory for all practical purposes.

The State of Massachusetts has always been the foremost state in the union for reliable and trustworthy statistical records, and in all probability it would be possible to secure the deaths by causes in 5-year age groups instead of ten-year groups. By taking the above table as a first approximation one should then obtain a very accurate table. On the other hand, it is possible to verify the final results in the above Life Table for Massachusetts by an entirely different process. It happens that the State of Massachusetts took a census in April 1915. This census for living males by attained ages could then be used as an approximation for the exposed to risk, while the deaths for the three years could be used as a basis for the number of deaths in a single year. A Life Table could then be constructed by means of the orthodox methods usually

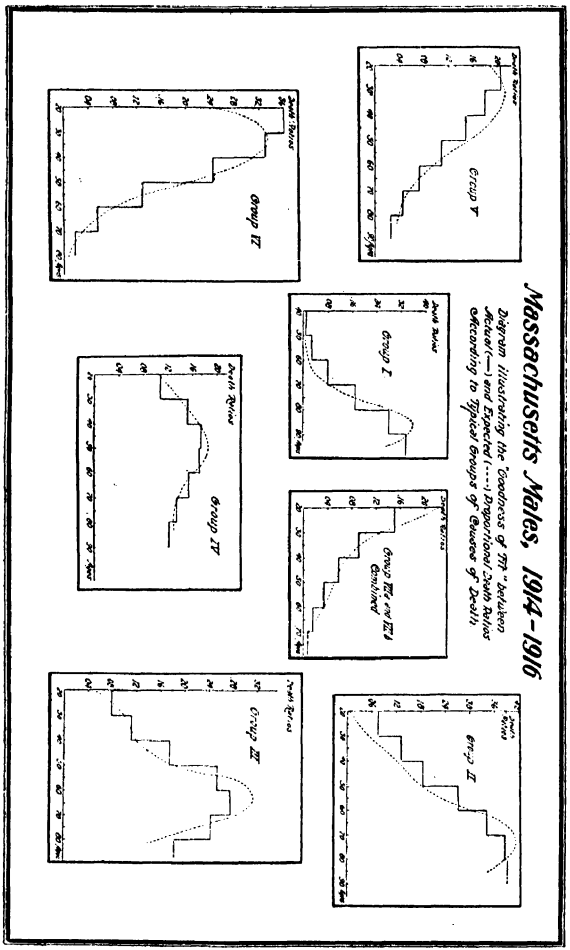
Massachusetts Males, 1904-1916
Diagram illustrating the "goodness of fit" between Actual (——) and Expected (·····) Proportional Death Rates according to Typical Groups of Causes of Death
Group V
Group II
Group I
Group IV
Group III
Group IV
Group III
Group II

Fig. 5.

Table I.

Expected or theoretical Proportionate Death Ratios as computed from the proposed system of frequency curves for Massachusetts Males, 1914—1916.

10			1.0	5.9	11.3	5.3	1.4	75.1
15			2.7	8.3	15.6	16 6	19.3	37.6
20		0.6	5.0	10.4	18.8	24.7	22.3	18.2
25		2.7	7.9	12.3	20.6	31.0	18.1	6.5
30		5.3	10.1	14.0	21.4	34.1	12.5	2.6
35		8.7	11.0	15.5	20.9	33.5	8.9	1.5
40		12.1	11.5	17.1	20.0	30.8	7.1	1.4
45		15.0	14.2	18.6	18.4	26.3	7.2	1.3
50	0.4	18.0	20.5	19.0	15.7	20.0	5.3	1.1
55	1.2	22.9	27.6	17.8	12.2	13.4	4.2	0.8
60	2.2	29.5	31.4	15.9	9.1	8.2	3.0	0.6
65	3.4	36.5	31.3	14.4	7.0	4.9	2.1	0.4
70	7.4	41.4	27.7	13.5	5.5	2.8	1.4	0.3
75	16.8	41.6	21.6	12.8	4.5	1.5	0.9	0.2
80	28.4	38.0	15.4	12.8	3.8	0.8	0.6	0.2
85	35.6	34.3	10.8	11.6	3.8	0.3	0.4	0.1
90	34.9	32.1	7.8	20.5	4.3	0.0	0.3	0.0

employed by actuaries and statisticians in the construction of mortality tables from census returns.

12. AMERICAN LOCOMOTIVE ENGINEERS' TABLE 1913—17. OTHER TABLES

As a third illustration, I shall construct a table for American Locomotive Engineers for the period 1913—1917. The statistical data forming the basic table are the mortuary records by attained age and cause of death among the members of The Locomotive Engineers' Life and Accident Insurance Association, a large fraternal order of the American Locomotive Engineers. The total number of deaths in the five year period amounted to more than 4,000. Distributed into separate groups of causes of death, it was found that it was possible to use a system of frequency curves similar to that employed in the State of Massachusetts, except for Group No. IV, for which it was found exceedingly difficult to find a single curve which would fit the data, and much points towards the actual presence of a compound curve of that group of causes of death among the Locomotive Engineers. The grouping of causes of death is, also slightly, different from that of Michigan and Massachusetts. I shall not go into further details as to the actual construction of this table, except to mention the areas of the various component fre-

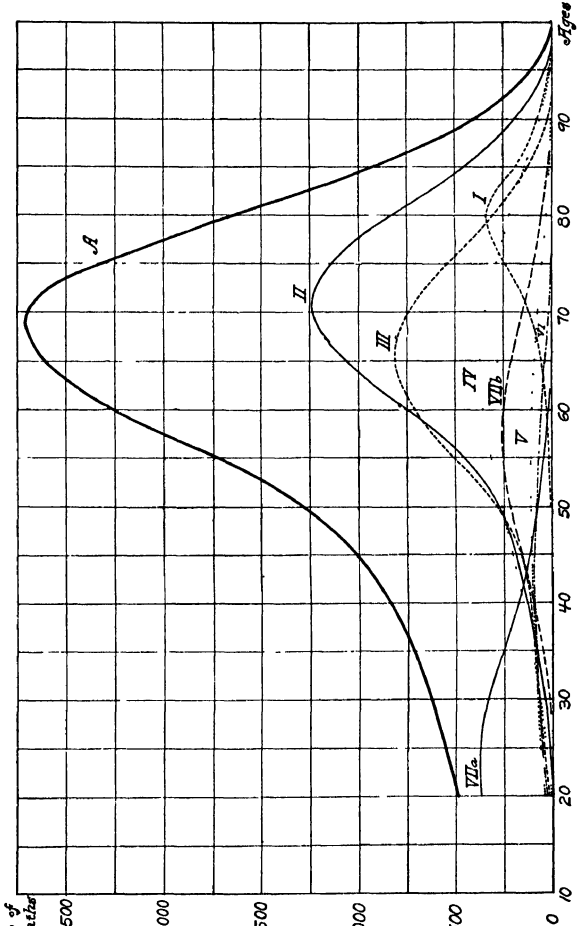

Fig. 6. American Locomotive Engineers.

quency curves of which I present the following table.

	Areas
Group I	44,857
— II	342,645
— III	226,022
— IV	147,420
— V	47,650
— VI	31,260
— VIIa	79,005
— VIIb	77,713
— VIII	3,428
	1,000,000

It must also be remembered that the radix of this table is taken at age 20, instead of at age 10 as is the case in the preceding tables. The final graph is shown on the preceding page. A number of diagrams illustrating the "goodness of fit" are also attached and need no further comment. It might, however, be of interest to mention the fact that the American actuary, Moir, has recently constructed a mortality table for American Locomotive Engineers along the orthodox lines from the data contained in the Medico-Actuarial Mortality investigation. Moir's table -- or at least the great bulk of the material from

which it was derived — falls in the interval between 1900 and 1913. Owing to the energetic "safety first" movement which since 1912 has been actively pursued by most of the leading American

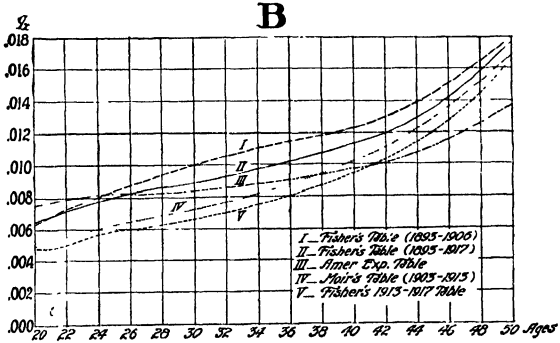

Fig. 7.

railroads, it is, however, to be expected that the period 1913—1917 indicates a reduced mortality as compared with that of Moir's period. This fact is also shown in the diagrams in Fig. 7.[1] On the other hand, the almost parallel movements of Moir's table with that of the table of the frequency curve method of 1913—1917, seems to indicate the soundness of the proposed method.

[1] Curves I, II and V are Locomotive Engineers' Mortality Tables for various periods.

12a. ADDITIONAL MORTALITY TABLES

A similar table showing mortality conditions among a decidedly industrial or occupational group has been constructed for coal miners in the United States. The original data of the deaths by ages and specific causes were obtained from the records of several fraternal orders and a large industrial life assurance company and comprised nearly 1600 deaths. The number of deaths above the age of sixty were, however, too few in number to determine with any degree of exactitude the area of component curves for the older age groups. For ages below sixty-five the table should on the other hand give a true representation of the mortality among coal miners in American collieries during the period under consideration[1]). A particular feature of this table is the comparatively low mortality in group VI, which contains primarily deaths from tuberculosis. Coal miners present in this respect different conditions than those usually prevailing in dusty trades where the death rate from tuberculosis is unusually high. The same feature is also borne out in previous investigations on the death rate of coal miners in Eng-

[1] It was not possible to seperate anthracite and bituminous coal miners. The data indicate, that anthracite mine workers have a higher accident rate than workers in bituminous mines.

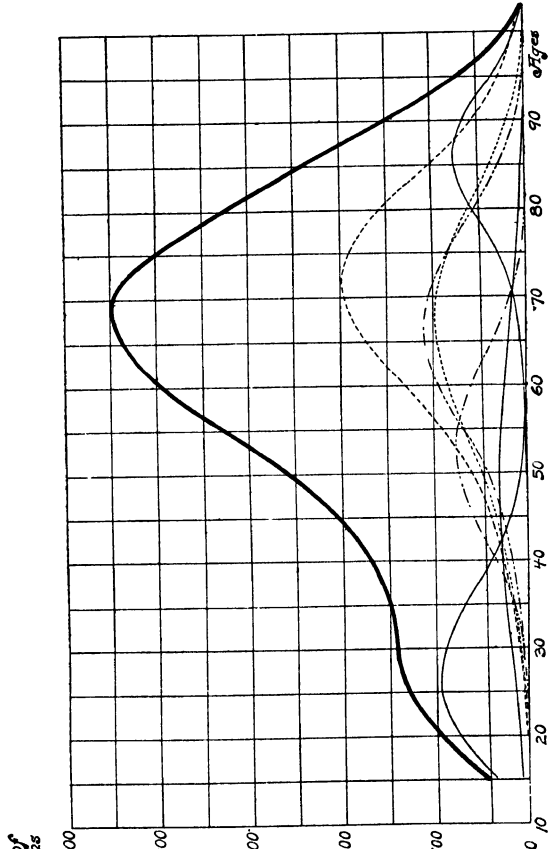

Fig. 8. American Coal Miners 1913—1917.

land, and by the recent investigations by Mr. F. L. Hoffman on dusty trades in America.

In order to have a measure of the mortality prevailing among industrial workers in America, we submit a table derived from a very detailed collection of mortuary records by age, sex and cause of death as published by the Metropolitan Life Insurance Company of New York. A deplorable defect in this splendid collection of data is the grouping together of all ages above seventy in a single age group, which makes it almost impossible to determine the component curves for higher ages with any degree of trustworthiness.

The defect in the original Metropolitan data for older age groups made it neccessary to modify the earlier sets or families of curves which were used on the Michigan and Massachusetts data and to combine several of the subsidiary component curves, especially those for the older age groups. Such modifications were, however, easily performed by means of simple logarithmic transformations.

I give below my grouping scheme for the Metropolitan data designated by the code numbers of the international list of causes of death. The actual cause of death corresponding to each code number is found under paragraph 8 of the present chapter.

GROUP I

10, 39 to 46, 48, 50, 54, 63 b, 64 to 66, 68, 79, 81, 82, 89 to 91, 94, 96, 97, 103, 105, 109 a, 120, 123, 124, 126, 127, 142, 154.

GROUP II

4, 13, 14, 18, 26, 27, 32 to 35, 47 (over age 20), 49, 51 to 53, 55, 60, 62, 70 to 72, 77, 78, 80, 83 to 88, 92, 95, 98 to 102, 106, 107, 109 b, 110 to 119, 122, 125, 143 to 145, 148, 149, 155 to 163.

GROUP III

28, 29, 31, 37, 38, 56 to 59, 67.

GROUP IV a AND IV b

1, 5 to 9, 17, 19, 20 to 25, 30, 61, 63 a, 73 to 76, 108, 146, 147, 150, 164 to 186, 47 (under age 20).

It will be noted that under this scheme Group I includes practically Groups I to III of the Michigan classification, Group II corresponds partly to IV and V for Michigan, Group III is practically Michigan's Group VI, while Group IV a and IV b takes in partly V, VII, and VIII in the Michigan experience. As a further correction I found it also advisable to transfer some of the deaths in the age intervals 10—14, 15—19, 20—24, and 25—29 in Groups I and II to Group IV a so as to avoid the long left tail ends in these older age curves.

After grouping the deaths (more than 200,000) of the Metropolitan experience according to the above scheme, it is a simple matter to compute the various

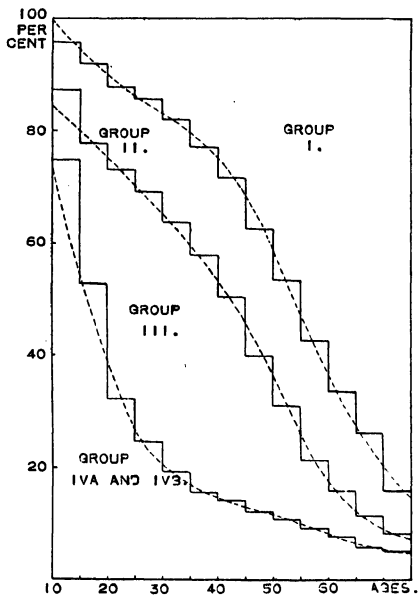

Fig. 9.

values of R(x) of the four groups for quinquennial age intervals and use these values (altogether 52 in number) for finding the observation equations and in the subsequent determination of the component curves as shown in the final mortality table in the appendix

to this chapter. A comparison between the observed values of R(x) by quinquennial ages and the continuous values of R(x) (indicated by dotted curves) as computed from the final mortality table is shown in Fig. 9. The "fit" between calculated and observed values is evidently satisfactory.

A most instructive and unique experience is offered in the table of Japanese Assured Males for the four year period 1914-1917 and based upon the death records of more than a dozen of the leading Japanese Life Assurance Companies. About 35,000 deaths by cause and arranged in quinquennial age groups were available for this construction. The component curves for the older age groups were determined by a simple logarithmic transformation of the variates and offered no particular obstacles in the a priori determination of the parameters. The curves for middle and younger life were more difficult to handle, especially the curves typical of tuberculosis, spinal meningitis and the peculiar Oriental disease known as Kakke, arising from an excessive rice diet. A first attempt to use the same curve types as employed in some of the European and American data did result in a very poor fit between the observed and calculated values of $R(x)$ for the younger age intervals clearly indicating that the clustering tendencies were different in the case of the Japanese data than in the other experiences I had previously dealt with.

The peculiar form of the observed values of $R(x)$ for the tuberculosis group indicated beyond doubt that the frequency curve for this group itself was a compound curve. I therefore decided to include both spinal meningitis and kakke with the tuberculosis group, and treat this new group as a compound frequency curve with two components. By successive trials I finally succeeded in establishing a complete curve system which satisfied the ultimate requirement of the fit between the observed and calculated values of $R(x)$ for the various groups.[1]

Grouping of Causes of Death in Japanese Assured Males 1914—1917.

GROUP I

Diseases of Arteries, Senility, Influenza, Cerebral Hemorrhage, Acute and Chronic Bronchitis, Bronchopneumonia.

GROUP II

Asthma and Pulmonary Emphysema, Cancer (all forms), Tumor, Diabetes, Other Diseases of Body, Paralytic Dementia, Tabes Dorsalis, Diseases of other organs for circulation of Blood, Chronic Nephritis, Other Diseases of Urinary Organs.

GROUP III

Mental Diseases, Other diseases of Spine and Medulla Oblongata, Other Diseases of Nervous

[1] See Addenda for the final table.

System, Diseases of Cardiac Valves, Pneumonia, Pleurisy, Other Respiratory Diseases, Gastric Catarrh, Ulcer of Stomach, Hernia, Other Diseases of Stomach, Diseases of Liver, Acute Nephritis, Diseases of Skin and Diseases of Motor Organs.

GROUP IV a AND IV b

Typhoid Fever, Malaria, Cholera, Acute Infectious Diseases, Peritonitis, Suicide, Dysentery, Tuberculosis (all forms), Syphilis, Kakke, Menengitis, Inflammation of the Caesum, Death by external causes (accidents, etc.).

Arranging the collected Japanese statistics on causes of death among assured males by attained age at death in accordance with the above scheme of grouping, using a 5 year interval as the unit, we obtain the following double entry table for the 35207 deaths as used in my computation for the various values of $R(x)$.

Ages	Group I	Group II	Group III	Group IV	Total
10—14	3	4	37	79	123
15—19	17	23	216	714	970
20—24	37	65	181	1640	1923
25—29	62	109	324	1975	2470
30—34	124	257	800	1993	3174
35—39	278	480	1147	2065	3970
40—44	449	662	1299	1674	4084
45—49	701	957	1352	1482	4491
50—54	742	959	1115	990	3806

Ages	Group I	Group II	Group III	Group IV	Total
55—59	864	1045	1041	728	3678
60—64	865	847	874	482	3068
65—69	626	571	612	186	1995
70—74	399	268	347	80	1094
75—79	123	76	100	20	319
80—84	16	13	10	3	42

The observed values of $R(x)$ as derived from the above table are shown in the staircase shaped histograph in Fig. 10. The correlated values of $R(x)$ as calculated from the final mortality table are shown as dotted curves on the same diagram. The "fit" between observed and calculated values of $R(x)$ is evidently satisfactory except for the youngest age intervals.

The construction of the present Japanese table constitutes probably the most severe trial to which the proposed method has hitherto been put. We are here dealing with an entirely different race living under different economic conditions than the nations of Europe and America and afflicted with certain forms of diseases which are comparatively rare or unknown among the Western nations.

It is therefore gratifying to note that the eminent Japanese actuary, Mr. T. Yano, in comparing the above mentioned table with an investigation he made on the aggregate mortality in 1913-1917 of all the Japanese life assurance companies (about 45 in number) from the actual number of lives exposed to risk

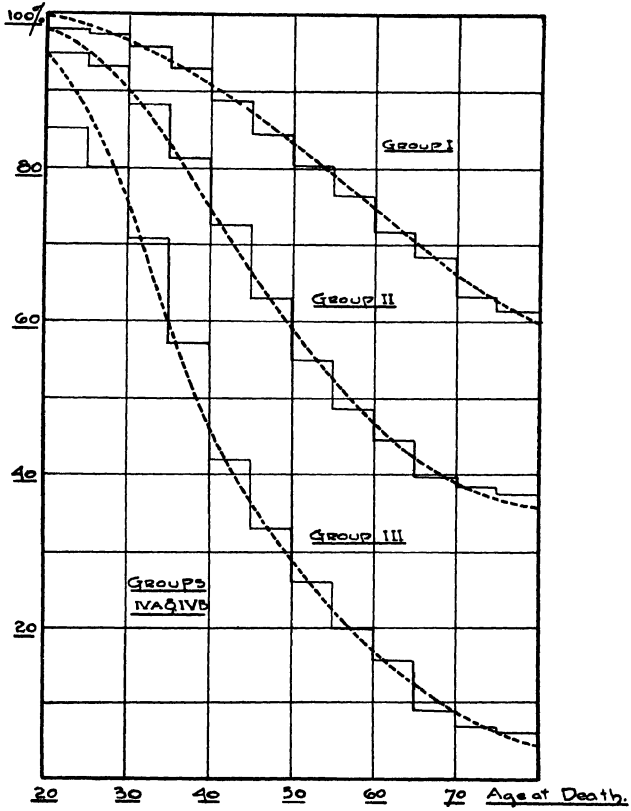
100%
80
60
40
20
GROUP I
GROUP II
GROUP III
GROUPS
IVA&IVB
20
30
40
50
60
70
Age at Death.

Fig. 10.

at various ages has been able to test independently the validity of the proposed method to complete satisfaction. (See remarks in preface).

13. CRITICISMS AND SUMMARY With these remarks I shall close the mere technical discussion of the proposed method and turn my attention to the arguments advanced by certain American critics against the possibility of constructing mortality tables from records of death alone. I deem no apology necessary to meet those critics and give a brief historical sketch of the origin of the proposed method, because remarks along this line will tend to accentuate the difficulties the mathematically trained biometrician has to contend with in obtaining a hearing among the present day school of actuaries and statisticians.

A good many critics, among whom I may mention Mr. John S. Thompson and Mr. J. P. Little, apparently have received an erroneous impression of the fundamental processes of the proposed method and its evident departure from the conventional methods. Mr. Thompson states "If we understand the process, the result is simply a graduation of "d_x" the "actual" deaths, and it is not apparent why a mortality table should not be formed from the unadjusted deaths and some other

function of graduation with equally good results"[1]. From this it would appear that Mr. Thompson is of the opinion that I have graduated the deaths as actually observed. As any one who will take the trouble to read the above article can see this is not the case. The actually observed numbers of deaths have only been used to construct the observed proportionate death ratios[2].

The whole process may be summarized as follows:

1) The choice (a priori) of a system of frequency curves based upon the hypothesis that the distribution of deaths according to age from typical causes of death can be made to conform to those postulated frequency curves whose parameters are known or chosen beforehand.

2) The grouping of causes of death so as to conform with the above mentioned system of frequency curves.

3) The computation for each age or age group of the proportionate death ratios of such groups

---

[1] Proceedings of the Casualty Actuarial Statistical Society of America, Vol. IV, Pages 399—400.

[2] These objections by Thompson and Little are shown in their full obscurity in the case of the tables for Locomotive Engineers, Coal Miners and Japanese Assured Males where the greatest number of observed deaths fell between ages 35—49.

from the collected statistical data of deaths by age and by cause of death.

4) The choice of approximate values of the areas of the various component frequency curves. Such approximate values can be determined by inspection or by simple linear correlation methods.

5) The determination by means of the theory of least squares of the various correction factors α with which the approximate values of the areas must be multiplied in order that we may obtain the probable values of the areas of the component curves. The observation equations necessary for this computation are obtained from the observed proportionate death ratios, which are independent of the exposed to risk.

6) The subsequent calculation of the products $NF(x)$ for all groups and for all integral ages. This gives us again the total number dying from all causes at integral ages among the original cohort of 1,000,000 entrants at age 10. In other words the d_x column from which the final mortality table can be constructed.

7) The computation of the "expected" or theoretical proportionate death ratios from the final table and their subsequent comparison with the "actual" or observed proportionate death ratios to illustrate the "goodness of fit".

It is this last step which constitutes the verifica-

tion of the results derived by means of a purely deductive or mathematical process, and is a test of very stringent requirements. It is namely required that there must be a *simultaneous* "fit", not alone for all groups of causes of death, but for all age intervals as well.

The sole justification of the proposed method hinges indeed upon the validity of the hypothesis. Is it indeed possible to choose a priori a system of frequency curves to which to fit our observed data? Theoretically speaking each population or sample population, as for instance certain occupational groups such as locomotive engineers, farmers, textile workers, miners, etc. will in all probability have its own particular system of frequency curves. From a purely practical point of view — and this is the one in which we are chiefly interested — we may, however, easily get along with a limited system af frequency curves for the various groups of causes of death and limit ourselves to a comparatively few sets of frequency curves to which to fit our statistical data. The case is analogous to that confronting a manufacturer of shoes. Undoubtedly the foot of one individual is different in form from that of any other individual, and in order to get an absolutely faultlessly fitting boot we would all have to go to a custom boot maker. Practical experience shows,

however, that it is possible to manufacture a few sizes of boots, say 6's, 7's, 8's and intermediate sizes in quarters and halfs, so as to fit to complete satisfaction the footwear of millions of people. Exactly in the same manner I have found from a long and varied experience in practical curve fitting that it is possible to fit the mortuary records of male deaths by attained age and cause of death to a comparatively limited number of sets of component curves, say not more than 5 or 6 sets. Moreover, if in a certain sample population a certain curve should not exhibit a satisfactory fit it is indeed a simple matter to change its parameters so as to improve the fit.

14. ADDITIONAL REMARKS ON PRINCIPLES OF METHOD

In regard to the classification of the causes of death into a limited number of groups it seems that some of the critics of the method are of the opinion that this classification is ironclad and fixed. This, however, is not the case. While in a specific sample population a certain cause of death might fall in group II, it is quite likely that the same cause of death would come under another group in another sample population. For instance, the deaths from asthma are in Michigan grouped under Group II. In the case of Coal Miners such deaths would, however, go into group

IV or group V. If the classification of causes of death were fixed, the frequency curves for separate population would show great variations, and it would be out of the question to limit ourselves to a small set of systems of component curves. Making the classification flexible, we are, on the other hand, in a better position to proceed with a fewer number of curves. For instance, in order to use the postulated frequency curve for Group VI for Michigan it was necessary to place the cause of death listed as No. 186 (other accidental traumatism) of the International Classification of Causes of Death in that group instead of in group V or VII, where most deaths of this type are ordinarily classed.

It would be interesting to see to what extent the proposed classification and the chosen system of frequency curves in Michigan deviates from the theoretically exact system of frequency curves. In the case of Michigan it would be impossible to test this. An approximate test might be obtained from the Michigan mortality data for the three year period 1909—1911. Professor Glover has constructed a mortality table for males in the State of Michigan in this three-year period by means of the usual methods employed by actuaries by resorting to the exposed to risk. Starting with a radix af 1,000,000 at age 10 it is possible to break

up the deaths or the d_x column of the Glover table into a set of subsidiary columns of death from groups of causes of death in the same order as given in Table A on page 133 by means of a simple application of the observed proportionate mortality ratios as derived from the 1909—1911 period. On the basis of a radix of 1,000,000 survivors at age 10 we find that according to the Glover Table, 5016 will die in the interval from 50—54. Let us also suppose that the proportionate mortality ratios in group III for ages 50—54 amounted to 0.23, then the number of deaths from group III in that particular interval in the Glover table would be $5016 \times 0.23 = 1154$. Similar numbers could be found for the other groups and for arbitrary age intervals, and we would in this manner have an empirical representation of the frequency curves. This aspect of the matter is treated in brief form on another page.

Returning now to our original discussion, it will readily be admitted that the method of constructing mortality tables by means of compound frequency curves cannot be considered as absolutely rigorous from the standpoint of pure mathematics. But neither can the usual methods of constructing mortality tables by graduation processes either by analytical formulas, mechanical interpolation formulas or a simple graphical process be considered

as mathematically exact. All statistical methods are, in fact, approximation processes. In the greater part of the realm of applied mathematics we have to resort to such approximation processes. It is thus absolutely impossible to solve correctly by ordinary algebraic processes simple equations of higher degree than the fourth. We encounter, however, in every day practice innumerable instances in which an approximation process, as for instance Newton's or Horner's methods or the method of finite differences, is sufficiently close to determine the roots of any equation so as to satisfy all practical requirements.

From this point of view I claim that the proposed method in the hands of adequately trained statisticians will yield satisfactory results, and I am inclined to think that the results are probably as true as the ones obtained by means of the usual methods, which especially in the case of graduation by interpolation formulas often are affected with serious systematic errors. Moreover, there are sound philosophical and biological principles underlying the proposed method, which is perhaps more than can be said about the usual methods, purely empirical in scope and principle. On the other hand, I will readily admit that the proposed method is by no means a simple rule of the thumb and it can under no circumstances be entrusted to

the hands of amateurs. The whole process can in my opinion only be employed when placed in the hands of the adequately trained statistician who is thoroughly familiar with his mathematical tools, as provided in the formulas from the probability calculus. Such adequate training is not acquired over night, but only through a long and patient study. Meticulous and patient work is often required before one is finally brought upon the right track, especially in the classification of the causes of death. Failure upon failure is oftentimes encountered by the beginner in this work, and it is probably only through such failures that the investigator is enabled to avoid the pitfalls of the often treacherous facts as disclosed by statistical data and steer a clear course. Mathematical skill is only acquired through a long and careful study. The illustrious saying of the Greek geometer, Euclid, who once told the Ptolemaian emperor that "there is no royal road in mathematics" holds true to-day as it did in the days of antiquity.

The fact that the method is no simple mechanical rule, but one which can be entrusted into skillful hands only, is, moreover, in my opinion, one of its strong points, because it eliminates all attempts of dilletantes to make use of it. A large manufacturing plant would not, for instance, put an ordinary blacksmith or horseshoer to work on

making the fine tools for certain parts of automatic machinery employed in the manufacture of staple articles. Only the most skilled and highly trained tool makers are able to produce machine parts, which often require precision measurements running into one thousandth part of an inch. Nor would a large contracting firm dream of putting a backwoods carpenter in charge of the construction of a skyscraper. Yet, this case is absolutely analogous to that of letting the mere collector of crude statistical data make an analysis and draw conclusions from certain collected facts as expressed in statistical series of various sorts.

While some American critics to all appearances have misunderstood the principles underlying the method, several European reviewers of the short summary of the method as originally published in the "*Proceedings of the Casualty Actuarial and Statistical Society of America*" evidently have understood its fundamental principles completely. The European critics seem, however, to be of the opinion that there is a rather prohibitive amount of arithmetical work involved in the actual construction of the mortality table. Thus a review in the *Journal of the Royal Statistical Society* for May 1918 has this to say:

> "Mr. Fisher's object is to construct a life table, being given only the deaths at ages and

not the population at risk. The hypothesis employed is that the total frequency of deaths can be resolved into specific groups of deaths, the frequencies of which cluster around certain ages. The parameters of these sub-frequencies having been determined, the areas are deduced from a system of frequency curves of the form:

$$R_B(x) = \frac{N_B F_B(x)}{N_B F_B(x) + N_C F_C(x) + N_D F_D(x) \ldots}$$

where $R_B(x)$, the proportional mortality at age x of deaths due to causes in group B and $F_B(x)$, is obtained from the equation of the sub-frequency curve for cause B, while $N_B + N_C + N_D + \ldots + N_K = 1{,}000{,}000$. The values of $R(x)$ provide a system of observational equations from which (by least squares) the values of N_B, &c., can be obtained.

"Since particularly in industrial statistics, or in general statistical inquiries under war conditions it is easier to obtain accurate data of deaths at ages than of exposed to risk, the success of the method is encouraging. It is, however, to be noted that the amount of arithmetical work envolved is considerable. Quite apart from the determination of the parameters of the frequency curves, the formation and solution of the normal equations needed to compute the areas is a heavy piece of work. It would be of interest to see whether the resolution into but three components effected by Professor Karl Pearson in his well-known

essay published in the "Chances of Death" could be made to describe with sufficient accuracy an ordinary tabulation of deaths from age 10 onwards to lead to approximately correct results for life table purposes. The test should, of course, be made with mortality data derived from a population very far from being stationary and the deductions compared with the results of standard methods. The subject is one of peculiar interest at the present time."

From the above quotation it is evident that this English reviewer has a clear conception of the fundamental principles upon which the method is based. His criticism is mainly directed against the heavy piece of arithmetical work involved. This work can, however, not be compared with the much more difficult task of obtaining the exposed to risk at various ages, which under all circumstances would take much greater time and be infinitely more costly, in fact be absolutely prohibitive from a financial point of view. I wish in this connection to state that the whole arithmetical work involved in the construction of the Michigan table was done by two computers in less than 70 hours, while the corresponding table for Massachusetts took about 75 hours. I do not know if this can be called exactly prohibitive.

In regard to the remarks of my British critic

concerning the Pearsonian method I might add that in my first attempt of an analysis of mortality conditions along the lines as described above I tried to subdivide the causes of death into four groups. It was, however, found that this was not always sufficient to describe the frequency distribution of the number of deaths around certain ages. I doubt whether it is at all possible to describe the frequency distribution in the various subgroups by a system of normal curves, which, of course, would somewhat lessen the work. I have made attempts to do this, but so far I have not been successful except in a few cases.[1] It might be possible that we should succeed in this if we first set up a hypothetically determined curve of the numbers exposed to risk. Such a curve might, for instance, be a normal curve. Personally, I believe that little would be gained by such a procedure. More fruitful appears an analysis by means of correlation surfaces. The mortality table constructed by the process as I have described it constitutes in its final form a correlation surface, wherein the age at death and the group of causes of death are the independent variables, and the number of deaths at a certain age and from a

[1] See Addenda for the Metropolitan Table and the Japanese Table.

certain group af causes of death is the numerical value of the correlation function of the two variates. Provided one could obtain an exact equation of such a correlation surface, it would be a simple matter to construct a mortality table, and I hope that some statistician may in the future be induced to attempt a solution of the problem in this light.

15. ANOTHER APPLICATION OF THE FREQUENCY CURVE METHOD

Before closing the discussion of this subject we shall, however, give a brief description of another application of compound frequency curves in the construction of mortality tables. We have here reference to the use of skew frequency curves in the graduation of crude mortality rates as computed in the usual empirical manner as the ratio of deaths to the number of lives exposed to risk at various ages. On page 165 it was mentioned that the State of Massachusetts took a census in April 1915. This census together with the deaths for the triennial period from 1914—1916 makes it an easy matter to construct a mortality table in the conventional manner. Moreover, such a table can be compared with the previously constructed table from mortuary records by sex, age and cause of death only and shown in the appendix.

In this connection it might be worth mention-

ing that my first table for Massachusetts as constructed by compound frequency curves was prepared during the summer of 1918 and first pre-

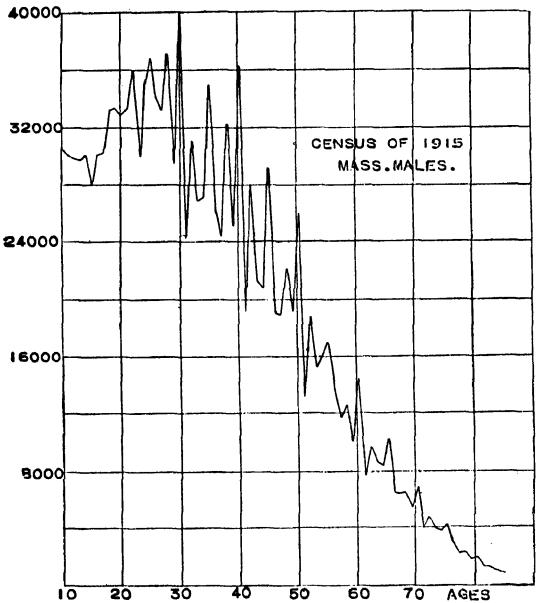

Fig. 11.

sented in a series of lectures delivered at the University of Michigan during the month of March 1919, while the final official report of the 1915 Massachusetts census did not come in the hands of the present writer before May 1919.

The official census of the population of Massachuetts by sex and single ages is given on page 478 in Vol. III of the Massachusetts report from which Fig. 11 has been constructed. It is seen from a mere glance of this graph that there is an unduly high tendency among the figures to cluster around ages being multiples of 5. This tendency is especially marked in the age interval 30—60 and presents a defect which is of no small importance in the construction of a mortality table by means of the conventional methods. It is indeed doubtful if a table constructed from data so greatly influenced by observation errors and misstatements of ages can be considered as absolutely trustworthy. On the other hand the data ought to be sufficiently exact to test the results arrived at by the proposed method of compound frequency curves.

We give below the male population in 5 year age groups for the middle census year of 1915 and the corresponding deaths from all causes during the triennial period 1914—1916.

MASSACHUSETTS

1915 Male Population and Number of Deaths among Males from 1914—1916.

Ages	Population, L_x.	Deaths 1914—16. D_x.
5— 9	169010	1715
10—14	152419	1004

Ages	Population, L_x.	Deaths 1914—16. D_x.
15—19	154773	1537
20—24	171961	2353
25—29	171017	2726
30—34	149294	2979
35—39	142617	3535
40—44	125462	4007
45—49	107909	4393
50—54	89490	5026
55—59	65133	5459
60—64	49079	5679
65—69	34790	6027
70—74	23638	5946
75—79	13724	4752
80—84	6494	3166
85—89	2479	1751
90—94	530	540
95—99	124	133
100 & over	12	23

A few small discrepancies will be found to exist between this table and the table printed on page 163, giving the observed deaths from various causes in ten year age intervals. This arises solely from the fact that a number of deaths were recorded where the contributing cause was unknown and could, therefore, not be distributed in their proper groups. But this defect is of no influence in the construction of mortality table by means

of the method of compound frequency curves, unless all the causes reported as unknown should happen to belong to the same group, which hardly can be assumed to be the case. At any rate the proportionate death ratios which are the keystone in this method of construction are for practical purposes left unaltered whether we include or exclude these few numbers of unknown causes. In the usual way of constructing tables from exposures and number of deaths it is on the other hand absolutely essential to include all deaths as otherwise the death rate will be underestimated.

Bearing these facts in mind we therefore refer to the above figures of L_x and D_x for Massachusetts Males from which we without further difficulty can construct an empirical mortality table, either by graphic methods or by simple summation or interpolation formulas. There is indeed no dearth of such formulas, of which a large number have been devised by Milne, Wittstein, Woolhouse, Higham, Sprague, Hardy, King, Spencer, Henderson, Westergaard, Gram, Karup and several other investigators. In the following computation I have used a formula originally devised by the Italian statistician, Novalis, and later on somewhat modified by the English actuary, King. The following schedule shows the actual process in detail.

MASSACHUSETTS MALES.

A. *Population.*

Graduated Quinquennial Pivotal Values.

Ages	Population L_x	$\triangle L_x$	$\triangle^2 L_x$	Age	Graduated Population
5— 9	169010	— 16591			
10—14	152419	+ 2354	+ 18945	12	29332
15—19	154773	+ 17188	+ 14834	17	30836
20—24	171961	— 944	— 18132	22	34537
25—29	171017	— 21723	— 20779	27	34369
30—34	149294	— 6677	+ 15047	32	29739
35—39	142617	— 17155	— 10478	37	28607
40—44	125462	— 17553	— 398	42	25095
45—49	107909	— 18419	— 866	47	21587
50—54	89490	— 24357	— 5938	52	17946
55—59	65133	— 16054	+ 8293	57	12961
60—64	49079	— 14289	+ 1765	62	9802
65—69	34790	— 11152	+ 3137	67	6933
70—74	23638	— 9914	+ 1238	72	4717
75—79	13724	— 8130	+ 1884	77	2731
80—84	6494	— 4015	+ 4115	82	1265
85—89	2479	— 1949	+ 2066	87	480
90—94	530	— 406	+ 1543	92	104
95—99	124	— 112	+ 294	97	23
100—104	12			102	1

Graduated Population $= u_{x+7} = 0.2 L_{x+5} - 0.008 \triangle^2 L_{x+5}$

B. *Deaths 1914—1916.*

Graduated Quinquennial Pivotal Values.

Ages	No. of Deaths D_x	$\triangle D_x$	$\triangle^2 D_x$	Age	Graduated Deaths
5— 9	1715	— 711			
10—14	1004	+ 533	+ 1244	12	200.8
15—19	1537	+ 816	+ 283	17	307.4
20—24	2353	+ 373	— 443	22	470.6
25—29	2726	+ 253	— 120	27	545.2
30—34	2979	+ 556	+ 303	32	595.8
35—39	3535	+ 472	— 84	37	707.0
40—44	4007	+ 386	— 86	42	801.4
45—49	4393	+ 633	+ 247	47	878.6
50—54	5026	+ 433	— 200	52	1005.2
55—59	5459	+ 220	— 213	57	1091.8
60—64	5679	+ 348	+ 128	62	1125.8
65—69	6027	— 81	— 429	67	1205.4
70—74	5946	— 1194	— 1113	72	1189.2
75—79	4752	— 1586	— 392	77	950.4
80—84	3166	— 1415	+ 171	82	633.2
85—89	1751	— 1211	+ 204	87	350.2
90—94	540	— 407	+ 804	92	108.0
95—99	133	— 110	+ 297	97	26.6
100—104	23			102	4.6

In this manner we obtain the graduated quinquennial pivotal values of the population and of the deaths for ages 12, 17, 22, 27, ... etc. Then

by dividing one third of the graduated deaths by the population we have the graduated pivotal values of the so-called "central death rates", or m_x for quinquennial ages from age 12 and up. From these values of m_x we easily find the corresponding values of q_x by means of the formula:

$$q_x = \frac{2\,m_x}{2 + m_x}$$

We give below the results of this computation

Massachusetts Males 1914—1916.

Age	1000 q_x from Novalis' Formula
12	2.21
17	3.33
22	4.64
27	5.29
32	6.68
37	8.25
42	10.65
47	13.53
52	18.67
57	26.38
62	38.29
67	58.12
72	81.90
77	109.91
82	165.02
87	240.18
92	325.64

The intervening values of q_x are without difficulty derived by interpolation formulas or by a graphical process. Once having all the values of q_x for separate ages from age 10 and up it is a simple matter to form tables of l_x and d_x commencing with a radix of 1,000,000 at age 10. Without going into tedious details we present the following values of l_x for decimal ages.

Massachusetts Males 1914—1916.

Age	l_x	Ages	Σd_x
10	1,000,000	10—19	27,700
20	972,300	20—29	47,330
30	924,970	30—39	66,750
40	858,220	40—49	98,650
50	759,570	50—59	153,900
60	605,670	60—69	233,150
70	372,520	70—79	237,130
80	135,390	80—89	124,760
90	10,640	90 & over	10,640
100	32		

16. GRADUATION OF d_x COLUMN BY FREQUENCY CURVES

It is to this table that we now shall apply a process of regraduation by means of the method of compound frequency curves. Here we have already an empirical representation of the total compound curve of death or the d_x curve.

This compound curve can now by simple and straightforward processes be broken up into its various component parts as to causes of deaths by means of the various observed proportionate mortality ratios, R_x shown in Table H on page 163.

Let us for the sake of illustration take the age interval 40—49. According to our empirically constructed table as derived from the Massachusetts 1915 census we find that the number of deaths among the survivors in this age interval amounts to 98,650.

Applying to this number the observed proportionate death ratios, R_x, in table H we are able to break this number up into its various component parts according to the groups of causes of death from which the numerical values of R_x were derived. These component parts are as follows:

Group	No. of Deaths
I	1180
II	18050
III	17170
IV	17170
V	14300
VI	23970
VII a & b	5820
VIII	990
	Total: 98650

Component Groups. Massachusetts Males 1914--1916.

Ages	I	II	III	IV	V	VI	VII	VIII	Total
10–14			730	690	1460	510	1360	6320	11100
15–19			1410	1210	3690	4150	2690	3450	16600
20–29		3600	3640	4870	9560	17800	6200	1660	47330
30–39		8610	7410	10210	11350	22490	5410	1270	66750
40–49	1180	18050	17170	17170	14300	23970	5820	990	98650
50–59	5230	41700	38630	26320	16620	19700	4920	780	153900
60–69	21450	79270	63420	35900	16780	12590	3500	240	233150
70–79	40070	91530	56910	31780	10670	4030	1900	240	237130
80–89	33560	49150	22460	14850	3120	1000	500	110	124750
90–	3590	4280	1230	1200	300	20	20	0	10640
	105080	296190	213010	144200	87850	106260	32350	15060	1000000

In the same manner we can break up the compound curve (the d_x curve) in its eight component parts for all other age intervals, which finally gives us the following table of component groups, printed on the preceeding page, and graphically this table will represent a series of frequency diagrams of the various groups of causes of deaths. It is an easy matter to fit such diagrams to a system of Laplacean-Charlier or Poisson-Charlier frequency curves, which symbolically may be represented as follows:

$$N_{\text{I}}F_{\text{I}}(x),\ N_{\text{II}}F_{\text{II}}(x)\ldots N_{\text{VIII}}F_{\text{VIII}}(x)$$

where $F(x)$ is the frequency function of the *percentage distribution* according to age of the various component groups or curves, while N stands for the areas of such curves.

These curve areas are simply the sub-totals of the respective groups in the above table. The parameters giving the equations of the curves $F_{\text{I}}(x)$, $F_{\text{II}}(x)$, $F_{\text{III}}(x)$, ... are easily computed by the methods of moments and are shown in the following table on page 207.

Once having determined the parameters of the various frequency curves it is a simple matter to construct the final mortality table which is shown in the addenda.

Values of Parameters of Component Curves, Massachusetts, 1914—1916 Males.[1]

Group	Mean	Dispersion	Skewness	Excess
I	75.0	9.78	+0.080	—0.005
II	67.5	13.65	+0.117	+0.017
III	64.0	14.12	+0.124	+0.030
IV	60.5	16.51	+0.089	—0.006
V	50.0	18.61	+0.026	—0.034
VI	43.5	15.57	—0.036	—0.023
VIIb	57.5	16.33	—0.027	—0.028

It now remains for us to compare the final values of q_x which we obtain from the three tables:

A) The values of q_x as computed in the usual

[1] In this grouping I have combined VII a and VIII into a single group and roughly fitted this group to a truncated Poisson-Charlier curve. This, of course, is not exact and introduces evidently errors in the younger age interval from 10—19. For ages above 20 this curve plays no importance and the other curves should for the ages above 20 give a satisfactory fit. If absolutely exactitude was required for younger ages it would indeed offer no difficulties to compute curves VIIa and VIII separately and thus obtain a much closer fit in the youngest age interval. In view of the fact that the present calculation is a test case only, it has not been thought necessary to go to these refinements. This defect will af course also effect to a slight extent group VII b.

way from the number of lives exposed to risk and the corresponding deaths at various ages.

B) The values of q_x as obtained by a re-graduation of the mortality table under A by means of compound frequency curves.

C) The values of q_x constructed from mortuary records by sex, age and cause of death, but without knowing the numbers of lives exposed to risk.

Massachusetts Males. 1914—1916.

Values of 1000 q_x by various methods.

Age	A	B	C
17	3.33	3.15	3.27
22	4.64	3.99	4.28
27	5.29	5.04	5.46
32	6.68	6.72	7.03
37	8.25	8.63	8.88
42	10.65	10.83	11.05
47	13.53	13.86	14.05
52	18.67	18.83	19.13
57	26.38	26.88	27.66
62	38.29	38.79	40.26
67	58.12	59.04	56.54
72	81.90	76.50	77.61
77	109.91	103.69	107.51
82	165.02	137.97	148.79

I think that every unbiased investigator will admit that there exists a close agreement be-

tween the three series. It is indeed difficult to say which one of the three is the most probable. We know that on account of the great perturbations due to misstatements of ages the values under A are effected with considerable errors. The usual interpolation or summation formulas do not suffice to remove these errors and tend often to increase them. A re-graduation by means of frequency curves as shown in series B will in all probability give better results, although on account of the large age interval (10 years) in which the causes of deaths are grouped in the Massachusetts reports this method does not come to its full right[1]. The values of q_x under A and B are naturally closely related to each other, and those in series B cannot be derived unless the values in series A are known beforehand. Series C on the other hand is independent of either A or B, having been derived by means of entirely different methods of construction.

17. COMPARISON BETWEEN DIFFERENT METHODS

A comparison between the parameters in the seperate component curves in *B* and *C* gives us, however, a way of testing the validity of the hypothesis upon which the method of

[1] See footnote on page 127.

series *C* rests. In the case of the series *C* we started with the hypothesis of the existence of a set of frequency curves of the *percentage distribution* of the number of deaths according to age among the various groups. On the basis of this hypothesis and from the observed values of the proportionate death ratios, R_x, we determined by the method of least squares the areas of this postulated set of frequency curves. In the case of the *B* series we broke up the empirically constructed compound death curve (the d_x curve) into its various component parts according to a similar classification of causes of deaths as under *C*. We have therefore in this case an empirical determination of the areas of the component curves and all that we need to do is to graduate the rough frequency diagrams as represented by such areas to a system of frequency curves.

Let us now briefly examine how far the various skew frequency curves in series *B* and *C* differ from each other. In regard to the various statistical parameters of the separate groups we have the following results:

Means.

Group	Series C	Series B
I	78.5	75.0
II	68.0	67.5

Group	Series C	Series B
III	63.0	64.0
IV	60.5	60.5
V	49.5	50.0
VI	44.0	43.5
VIIb	57.5	57.5

Dispersions.

Group	Series C	Series B
I	7.98	9.78
II	12.21	13.65
III	13.05	14.12
IV	17.86	16.51
V	18.51	18.61
VI	14.68	15.57
VIIb	12.16	16.33

Skewness.

Group	Series C	Series B
I	+0.092	+0.080
II	+0.115	+0.117
III	+0.121	+0.124
IV	+0.098	+0.089
V	+0.033	+0.026
VI	—0.010	—0.036
VIIb	—0.002	—0.027

	Excess.	
Group	Series C	Series B
I	—0.033	—0.005
II	+0.023	+0.017
III	+0.047	+0.030
IV	—0.009	—0.006
V	—0.031	—0.034
VI	—0.027	—0.023
VIIb	—0.003	—0.028

Taken all in all there is found to exist a satisfactory agreement between the hypothetical values in series *C* and the values derived by empirical methods. It is only in group *I* that we find some important discrepancies. This group contains causes of death typical of extreme old age where we naturally may expect great perturbations owing to large errors from random sampling, especially in series *B*. In this same connection we may also mention that the empirically determined values under series *B* are subject to a slight correction by means of the Sheperd formulas, which were not employed in my computations.

We have already mentioned that the system of frequency curves which we choose a priori for Massachusetts (Series C) was the same system which we had used on a previous occasion in the construction of a mortality table for Eng-

lish Males for the period 1911—1912[1]). This is a fact of no small importance. It will in general be found that the *percentage distribution* according to age in the various component curves differs little in different sample populations. Even in the case of American Locomotive Engineers it was found possible to use the same set of curves as in the case of Massachusetts and England and Wales. In the same way I have found that the set of curves used in the construction of the table of Michigan Males also can be used in the case of males in the urban population of Denmark. With a very few exceptions I have found it possible to get along with a limited number of sets of curves, say four or five sets. Should it nevertheless prove impossible to fit the original data to any one of these particular curve systems, it will in most cases be found possible by means of successive approximations to reach a system of curves which may be made the a priori basis for the construction of the final table as was the case in the table for Japanese assured males.

Finally we come to the comparison of the various areas of the component curves. We have here:

---

[1] See "Proceedings of the Casualty Actuarial Society of America", Vol. IV, page 409.

	Areas. C	B
I	90064	105000
II	281470	296190
III	207854	213010
IV	151316	144200
V	99543	87850
VI	107718	106260
VII & VIII	62035	47410
Total	1000000	1000000

Evidently the agreement is not so close in this case. But it would indeed be rather rash to assert that the values in series *C* are faulty. One must here bear in mind the diametrically opposite principles employed in the determination of these areas. In series *B* we have a direct determination by empirical methods. In this determination we shall, however, find reflected all the original systematic and observational errors originally present in series *A* from which the curves under *B* were computed. Every error due to misstatements of ages and systematic errors introduced by the summation or interpolation formulas will be directly reflected in the areas under series *B*, and such areas can therefore in a sense only be considered as a first approximation to the true or presumptive areas.

Another point well worth remembering is the one that no conditions are imposed upon the areas in series B. In series C where we work with mortuary records only we have on the other hand the very important condition or restriction requiring that the areas of the component curves must be so determined that their ratios to the compound curve for various age intervals will conform as closely as possible with the observed proportionate death ratios, R_x, for those same age intervals.

In order to test the influence of this additional requirement in respect to conformity to observed proportionate death ratios we might use the values of the component curves under series B as a first approximation and then afterwards determine the correction factors α for the areas in exactly the same way as in the case of series C. No doubt such a calculation would tend to improve the table.

A difficulty occurs, however, in the case of the Massachusetts data owing to the large interval of 10 years into which the causes of death by attained ages are grouped. As pointed out in the footnote on page 127 the quantity $R_B(x)$, (x = 10, 11, 12, ... 100; B = I, II, III, ...), can only be considered as being independent of the "exposed to risk" if the age interval into which the deaths fall is sufficiently small. If this is not

the case, the "central" values of $R_B(x)$ are subject to certain corrections. In the case of the groups of causes of death typical of younger ages the observed "central" values of $R_{VII}(x)$ and $R_{VIII}(x)$ for the age intervals 10—19, 20—29, 30—39 are evidently too high, while on the other hand the values of $R_I(x)$ and $R_{II}(x)$ in the case of the age intervals 60—69, 70—79, 80—89, 90—100 are too low as compared with the true values of $R(x)$ at these "central" ages. I have, however, tacitly ignored this fact in my computations. The subsequent result is that the final values of q_x for the younger ages in column C as shown on page 208 are in all probability a little too high, and the values of q_x above 65 too low. In the case of the other tables as shown in the present book the age interval into which the causes of death were arranged was 5 years or less, and the error was thus reduced to such an extent that further corrections may be disregarded for all practical purposes.

ADDENDA I

Showing Detailed Mortality Tables and Death Curves for

1) Japanese Assured Males (1914—1917)
2) Metropolitan Life. White Males (1911—1916)
3) American Coal Miners (1913—1917)
4) American Locomotive Engineers (1913—1917)
5) Massachusetts Males (Series C) (1914—1916)
6) Michigan Males (1909—1915)
7) Massachusetts Males (Series B) (1914—1916).

Mortality Table—Japanese Assured Males

1914—1917 (Aggregate Table)

Age	I	II	III	IVa	IVb	dx	lx	1000qx
15	24	65	343		2379	2811	1000000	2.81
16	39	74	360		3645	4118	997189	4.13
17	43	84	388		4888	5403	993071	5.44
18	48	93	415		5981	6557	987668	6.64
19	54	107	446		6826	7433	981111	7.58
20	60	120	478		7447	8105	973678	8.32
21	68	135	513		7716	8432	965573	8.73
22	77	153	550	12	7734	8526	957141	8.91
23	87	171	591	27	7581	8457	948615	8.92
24	101	195	633	50	7274	8253	940158	8.86
25	111	218	678	77	6864	7948	931905	8.53
26	126	246	729	112	6384	7597	923957	8.22
27	140	278	780	153	5860	7211	916360	7.87
28	160	315	838	206	5341	6860	909149	7.54
29	178	353	899	268	4821	6519	902289	7.22
30	198	395	963	341	4323	6220	895770	6.94
31	227	446	1033	425	3853	5984	889550	6.73
32	252	501	1109	521	3421	5804	883566	6.59
33	286	557	1185	629	3021	5678	877762	6.46
34	319	626	1273	751	2665	5633	872084	6.46
35	358	700	1364	885	2336	5643	866451	6.51
36	401	779	1460	1031	2048	5719	860808	6.64
37	450	872	1564	1186	1797	5869	855089	6.86
38	502	970	1671	1350	1566	6059	849220	7.13
39	570	1081	1791	1524	1366	6332	843161	7.51
40	638	1197	1916	1701	1191	6643	836829	7.94
41	716	1332	2049	1883	1037	7017	830186	8.45
42	802	1475	2193	2066	903	7439	823169	9.04
43	899	1632	2341	2249	783	7904	815730	9.69
44	1005	1799	2501	2428	680	8413	807826	10.41
45	1126	1985	2671	2599	598	8979	799413	11.23
46	1261	2180	2852	2764	514	9571	790434	12.10
47	1406	2393	3042	2917	447	10205	780863	13.07
48	1575	2611	3236	3061	395	10878	770658	14.12
49	1754	2867	3459	3187	339	11606	759780	15.27
50	1957	3122	3666	3298	295	12338	748174	16.49
51	2180	3395	3892	3389	257	13113	735836	17.82
52	2426	3679	4136	3473	224	13938	722723	19.29
53	2692	3984	4380	3532	195	14783	708785	20.86
54	2987	4285	4638	3576	172	15658	694002	22.56
55	3306	4610	4922	3611	147	16596	678344	24.47
56	3654	4940	5177	3612	130	17513	661748	26.46
57	4026	5274	5456	3605	113	18474	644235	28.68
58	4432	5603	5742	3581	97	19455	625761	31.09
59	4857	5937	6025	3544	84	20447	606306	33.72
60	5316	6257	6316	3498	74	21461	585859	36.63
61	5795	6568	6604	3424	69	22460	564398	39.79
62	6293	6860	6890	3345	59	23447	541938	43.27
63	6805	7129	7162	3255	51	24402	518491	47.15
64	7332	7361	7423	3150	43	25309	494089	51.22
65	7854	7570	7672	3042	38	26176	468780	55.84

Age	I	II	III	IVa	IVb	dx	lx	1000qx
66	8366	7727	7896	2919	36	26944	442604	60.88
67	8863	7838	8089	2791	31	27612	415660	66.43
68	9313	7894	8257	2655	28	28147	388048	72.53
69	9719	7894	8385	2511	23	28532	359901	79.27
70	10053	7829	8468	2362	20	28732	331369	86.71
71	10294	7700	8503	2212	18	28727	302637	94.92
72	10424	7496	8477	2067	15	28479	273910	103.97
73	10424	7227	8380	1901	13	27954	245431	110.69
74	10280	6897	8230	1746	13	27166	217477	124.91
75	9970	6503	8002	1593	10	26078	190311	137.02
76	9492	6057	7695	1444	10	24698	164233	150.38
77	8834	5571	7313	1298	8	23024	139535	165.00
78	8037	5047	6853	1159	7	21103	116511	181.12
79	7086	4499	6314	1026	6	18931	95408	198.42
80	6046	3943	5733	900	5	16621	76477	217.33
81	4953	3400	5091	784	4	14232	59856	237.77
82	3871	2862	4421	676	3	11833	45624	259.35
83	2813	2365	3730	577	2	9487	33791	280.75
84	1957	1907	3046	489	1	7400	24304	304.48
85	1232	1498	2396	412		5538	16904	327.61
86	701	1141	1797	340		3979	11366	350.08
87	343	844	1275	277		2739	7387	370.76
88	140	603	844	225		1812	4648	389.78
89	48	408	516	179		1151	2836	405.85
90	14	269	283	141		707	1685	419.58
91	5	171	134	110		420	978	429.44
92		111	53	83		247	558	442.65
93		56	14	63		133	311	452.10
94		28	4	44		76	178	457.05
95		14	2	31		47	102	460.78
96		5	1	22		28	55	509.01
97				14		14	27	518.50
98				9		9	13	692.30
99				4		4	4	1000.00

Mortality Table

Metropolitan White Males 1911—1916

Age	I	II	III	IVb	IVa	dx	lx	1000qx
10	80	153	205	47	1720	2205	1000000	2.21
11	95	179	274	61	1776	2385	997795	2.39
12	118	210	350	77	1812	2567	995410	2.58
13	141	244	444	96	1832	2757	992843	2.78
14	168	282	550	116	1834	2950	990086	2.98
15	202	327	671	140	1825	3165	987136	3.21
16	240	373	810	171	1803	3397	983971	3.45
17	282	427	960	199	1772	3640	980574	3.71
18	336	483	1130	233	1733	3915	976934	4.01
19	393	545	1315	274	1680	4207	973019	4.32
20	454	611	1514	311	1612	4502	968812	4.65
21	527	685	1728	358	1539	4837	964310	5.02
22	599	765	1951	407	1449	5169	959473	5.39
23	687	845	2184	459	1363	5538	954304	5.80

Age	I	II	III	IVb	IVa	dx	lx	1000qx
24	775	932	2428	515	1279	5929	948766	6.25
25	874	1024	2674	575	1190	6337	942837	6.72
26	977	1120	2924	638	1107	6766	936500	7.32
27	1088	1223	3173	703	1012	7199	929734	7.74
28	1202	1328	3414	770	923	7637	922535	8.28
29	1324	1436	3648	839	840	8087	914898	8.84
30	1473	1549	3879	909	757	8567	906811	9.45
31	1584	1662	4089	985	684	9004	898244	10.02
32	1702	1779	4283	1052	614	9430	889240	10.60
33	1863	1899	4459	1125	545	9891	879810	11.24
34	2012	2015	4604	1196	485	10312	869919	11.85
35	2160	2139	4740	1266	427	10732	859607	12.48
36	2324	2259	4842	1332	378	11135	848875	13.12
37	2485	2379	4919	1399	335	11517	837740	13.75
38	2664	2501	4968	1462	296	11891	826223	14.39
39	2847	2617	4989	1520	258	12231	814332	15.02
40	3057	2734	4988	1577	226	12578	802101	15.68
41	3272	2848	4953	1628	192	12893	789523	16.33
42	3508	2960	4898	1675	163	13204	776630	17.00
43	3767	3066	4821	1719	143	13516	763426	17.70
44	4057	3170	4719	1757	120	13823	749910	18.43
45	4389	3267	4604	1789	100	14149	736087	19.22
46	4748	3358	4471	1816	90	14483	721938	20.06
47	5153	3447	4320	1839	75	14834	707455	20.97
48	5599	3526	4160	1855	61	15201	692621	21.95
49	6064	3598	3991	1867	50	15590	677420	23.01
50	6631	3663	3810	1872	42	16018	661830	24.20
31	7198	3721	3630	1872	35	16456	645812	25.48
52	7820	3769	3443	1867	30	16929	629356	26.90
53	8492	3809	3254	1857	22	17434	612427	28.47
54	9168	3839	3069	1840	10	17926	594993	30.13
55	9897	3858	2876	1820	1	18452	577067	31.98
56	10637	3868	2696	1793		18994	558615	34.00
57	11378	3867	2519	1762		19526	539621	36.18
58	12114	3853	2340	1726		20033	520095	38.52
59	12847	3830	2169	1687		20533	500062	41.06
60	13555	3794	2004	1640		20591	479529	43.77
61	14217	3746	1844	1591		21396	358538	46.67
62	14817	3685	1692	1541		21735	437140	49.72
63	15359	3615	1547	1484		22005	415405	52.97
64	15820	3535	1408	1425		22188	393400	56.40
65	16179	3443	1277	1364		22263	371212	59.97
66	16450	3340	1153	1299		22242	348949	63.74
67	16610	3229	1037	1235		22111	326707	67.68
68	16691	3109	930	1166		21896	304596	71.89
69	16591	2981	828	1098		21498	282700	76.05
70	16412	2851	736	1030		21029	261202	80.51
71	16107	2711	649	955		20422	240173	85.03
72	15721	2568	571	892		19752	219751	89.88
73	15225	2423	500	825		18973	199999	94.87
74	14629	2271	434	759		18093	181026	99.95
75	13946	2126	377	695		17144	162933	105.22
76	13225	1976	325	632		16158	145789	110.83
77	12423	1828	278	572		15101	129631	116.49
78	11580	1684	237	515		14016	114530	122.38

Age	I	II	III	IVb	IVa	dx	lx	1000qx
79	10729	1543	200	461		12933	100514	128.67
80	9840	1406	167	411		11824	87581	135.01
81	8950	1272	138	363		10723	75757	141.54
82	8092	1144	115	318		9669	65034	148.68
83	7237	1024	98	282		8641	55365	156.07
84	6420	911	79	247		7657	46724	163.88
85	5645	806	65	208		6724	39067	172.11
86	4920	707	53	181		5861	32343	181.21
87	4240	615	43	150		5048	26482	190.62
88	3622	531	34	126		4313	21434	201.22
89	3065	457	27	106		3655	17121	213.48
90	2550	387	22	87		3046	13466	226.20
91	2099	327	16	70		2512	10420	241.07
92	1698	270	14	56		2038	7908	257.71
93	1355	222	11	45		1633	5870	278.19
94	1053	179	8	35		1275	4237	300.92
95	805	143	6	27		981	2962	331.20
96	595	112	5	20		732	1981	369.51
97	412	85	1	14		512	1249	409.93
98	286	62		10		358	737	485.75
99	198	27		6		231	379	609.50
100	95	15		4		114	148	770.27
101	27	5		2		34	34	1000.00

Mortality Table—American Coal Miners

(1913—1917)

Age	I	II	III	IV	Va	Vb	VI	dx	lx	1000qx
18		99	124	142	4566	7	366	5304	1000000	5.30
19		114	144	164	4702	10	408	5542	994696	5.57
20		140	168	187	4954	14	452	5915	989154	5.98
21		162	194	214	5196	19	498	6283	983239	6.39
22		190	223	243	5234	27	546	6463	976956	6.62
23		223	250	272	5151	38	597	6531	970493	6.73
24		256	282	307	5067	50	646	6608	963962	6.86
25		298	315	341	4952	69	697	6672	957354	6.97
26		341	349	379	4846	91	749	6755	950682	7.11
27		390	386	421	4748	120	802	6867	943927	7.27
28		440	424	465	4683	156	853	7021	937060	7.49
29		498	461	508	4569	202	903	7141	930039	7.68
30		557	500	560	4413	257	953	7240	922898	7.84
31		622	538	609	4220	326	1002	7317	915658	7.99
32		688	579	663	4000	408	1048	7386	908341	8.13
33		761	618	718	3757	505	1093	7452	900955	8.27
34		837	654	777	3500	618	1133	7519	893503	8.42
35		915	693	840	3233	749	1175	7605	885984	8.58
36		994	732	905	2963	898	1212	7704	878379	8.77
37		1084	775	973	2697	1064	1246	7839	870675	9.00
38		1171	818	1045	2435	1251	1277	7997	862836	9.27
39		1267	867	1124	2184	1452	1305	8199	854839	9.59
40		1364	920	1206	1946	1667	1329	8432	846640	9.96
41		1471	978	1293	1723	1894	1352	8711	838208	10.39
42		1581	1045	1386	1515	2131	1369	9027	829497	10.88

Age	I	II	III	IV	Va	Vb	VI	dx	lx	1000qx
43		1705	1125	1489	1325	2372	1383	9399	820470	11.46
44		1835	1222	1585	1106	2609	1395	9752	811071	12.02
45	1	1976	1322	1712	883	2841	1403	10133	801319	12.65
46	6	2132	1444	1837	853	3063	1408	10743	791186	13.58
47	10	2302	1584	1971	729	3265	1410	11271	780443	14.44
48	21	2492	1741	2114	619	3443	1408	11838	769172	15.39
49	32	2705	1918	2265	524	3595	1402	12441	757334	16.43
50	42	2934	2118	2423	442	3706	1395	13060	744893	17.53
51	54	3190	2337	2589	368	3790	1383	13711	731833	18.74
52	73	3470	2567	2764	307	3832	1368	14380	718122	20.02
53	94	3775	2820	2945	255	3832	1352	15073	703742	21.42
54	123	4104	3086	3130	210	3790	1331	15774	688669	22.91
55	153	4437	3355	3313	173	3706	1308	16445	672895	24.44
56	185	4843	3637	3501	141	3595	1281	17183	656450	26.18
57	225	5246	3922	3689	115	3443	1252	17892	639267	27.99
58	268	5656	4192	3872	93	3265	1220	18566	621375	29.88
59	310	6085	4454	4047	76	3063	1186	19221	602809	31.89
60	354	6530	4703	4209	61	2841	1148	19846	583588	34.01
61	402	6970	4936	4364	48	2609	1109	20438	563742	36.25
62	450	7403	5133	4500	39	2372	1076	20964	543304	38.59
63	508	7832	5305	4618	30	2131	1023	21447	522340	41.05
64	573	8230	5438	4718	24	1894	978	21855	500893	43.63
65	648	8615	5533	4795	19	1667	931	22208	479038	46.36
66	746	8954	5581	4846	15	1452	884	22478	456830	49.20
67	875	9255	5596	4871	13	1251	834	22695	434352	52.25
68	1015	9507	5563	4871	9	1064	785	22814	411657	55.41
69	1207	9704	5479	4841	6	898	736	22871	388843	58.81
70	1437	9846	5358	4786	6	749	686	22868	365972	62.49
71	1702	9917	5196	4701	4	618	637	22775	343104	66.38
72	2008	9931	4999	4592	4	505	588	22627	320329	70.64
73	2334	9871	4771	4460	2	408	540	22386	297702	75.20
74	2677	9747	4513	4302	2	326	494	22061	275316	80.10
75	3028	9557	4233	4125	2	257	449	21651	253255	85.49
76	3332	9307	3941	3929	1	202	408	21120	231604	91.19
77	3610	9001	3638	3722	1	156	366	20494	210484	97.37
78	3827	8643	3322	3496		120	329	19737	189990	103.88
79	3967	8237	3012	3267		91	293	18867	170253	110.82
80	4020	7799	2704	3029		69	258	17879	151386	118.10
81	3980	7327	2411	2788		50	226	16782	133507	125.70
82	3916	6803	2123	2552		38	198	15630	116725	133.90
83	3658	6315	1846	2313		27	171	14330	101095	141.75
84	3370	5801	1596	2085		19	147	13018	86765	150.04
85	3040	5286	1366	1862		14	125	11693	73747	158.56
86	2684	4776	1151	1650		10	105	10376	62054	167.21
87	2305	4281	957	1448		7	88	9086	51678	175.82
88	1937	3809	789	1261		5	71	7872	42592	184.82
89	1584	3353	640	1085		3	60	6725	34720	193.69
90	1269	2924	513	927		2	48	5683	27995	203,00
91	985	2535	404	784		2	38	4748	22312	212.80
92	747	2168	310	650		1	29	3905	17564	222.33
94	551	1845	231	531			22	3180	13659	232.81
94	396	1545	170	428			17	2556	10479	243.92
95	278	1279	119	338			12	2026	7923	255.71
96	198	1050	79	261			7	1594	5897	270.31
97	126	845	48	195			5	1219	4303	283.29

Age	I	II	III	IV	Va	Vb	VI	dx	lx	1000qx
98	85	672	26	140			2	925	3084	299.94
99	70	525	9	96				701	2159	324.69
100	35	401		59				495	1458	339.51
101	24	298		29				351	963	364.48
102	19	217		4				240	612	392.16
103	14	149						163	372	438.17
104	10	97						107	209	511.96
105	8	55						63	102	727.65
106	6	25						37	39	794.87
107	3	2						5	8	625.00
108	2							2	3	666.67
109	1							1	1	1000.00

American Locomotive Engineers' Mortality Table
(1913—1917)

Age	I	II	III	IV	V	VI	VIIa	VIIb	VIIIa	VIIIb	dx	lx	1000qx
20		25	206	189	349	280	3688	22	105	13	4877	1000000	4.88
21		64	239	218	377	310	3866	29	97	14	5224	995123	5.25
22		78	277	249	404	342	3938	37	88	17	5430	989899	5.49
23		97	318	281	432	375	3894	43	75	19	5534	984469	5.62
24		123	360	310	460	408	3801	58	62	23	5605	978935	5.73
25		154	405	387	488	443	3669	72	51	25	5694	973330	5.85
26		188	453	404	517	478	3571	88	39	27	5765	967636	5.96
27		224	505	453	544	510	3454	117	30	29	5866	961871	6.10
28		267	554	502	574	544	3324	138	25	32	5960	956005	6.23
29		319	602	559	602	576	3183	167	19	33	6060	950045	6.38
30		379	651	620	630	607	3031	204	14	38	6174	943985	6.54
31		440	699	680	655	637	2894	248	10	39	6302	937811	6.72
32		513	744	749	682	664	2758	290	9	42	6451	931509	6.93
33		592	788	821	707	691	2628	349	6	43	6625	925058	7.16
34		673	830	897	734	715	2504	408	4	46	6811	918433	7.42
35		766	870	977	758	736	2382	474	1	48	7012	911622	7.70
36		855	909	1060	781	756	2257	553	1	51	7223	904610	7.98
37		952	947	1144	804	772	2134	634		52	7439	897387	8.29
38		1059	990	1237	826	787	2016	728		53	7696	889948	8.65
39		1162	1034	1333	843	798	1890	823		56	7939	882252	9.00
40		1273	1089	1431	865	808	1769	932		58	8225	874313	9.41
41		1383	1154	1530	883	814	1652	1041		59	8516	866088	9.83
42		1498	1229	1635	901	818	1539	1150		61	8831	857572	10.30
43		1618	1322	1745	916	820	1428	1275		62	9186	848741	10.82
44		1736	1440	1851	932	818	1321	1397		62	9567	839555	11.40
45		1869	1584	1963	946	816	1222	1521		64	9985	829988	12.03
46	6	2017	1754	2078	957	810	1133	1645		65	10465	820003	12.76
47	10	2168	1964	2195	968	801	1036	1769		65	10976	809538	13.56
48	13	2342	2212	2309	977	792	942	1893		65	11545	798562	14.46
49	18	2544	2497	2425	984	779	859	2001		65	12172	787017	15.47
50	26	2768	2813	2541	991	764	768	2112		65	12848	774845	16.58
51	34	3031	3166	2658	995	746	686	2213		65	13594	761997	17.84
52	46	3338	3556	2769	997	728	610	2300		65	14409	748403	19.25
53	59	3679	3979	2880	997	706	535	2372		65	15272	733994	20.81
54	77	4072	4419	2989	996	685	461	2439		64	16202	718722	22.54
55	96	4517	4878	3089	992	659	409	2489		64	17193	702520	24.47
56	116	5011	5338	3215	986	632	357	2519		62	18236	685327	26.61

57	141	5549	5797	3280	979	605	318	2534	61	19264	667091	28.88
58	168	6127	6247	3362	969	576	280	2534	61	20324	647827	31.37
59	194	6749	6677	3441	957	545	235	2511	59	21368	627503	34.05
60	222	7401	7060	3505	942	514	197	2482	58	22381	606135	36.92
61	252	8107	7407	3557	926	483	159	2431	56	23378	583754	40.04
62	282	8750	7695	3600	905	451	99	2366	53	24201	560376	43.19
63	318	9392	7922	3633	884	418	52	2286	52	24957	536175	46.55
64	359	10026	8079	3649	861	387	24	2198	49	25632	511218	50.14
65	406	10615	8161	3649	836	355	7	2096	48	26173	485586	53.90
66	467	11146	8170	3640	809	324		1987	45	26588	459413	57.87
67	548	11605	8096	3616	778	293		1864	43	26842	432825	62.02
68	636	11971	7953	3574	749	265		1755	40	26943	405982	66.37
69	756	12238	7737	3518	717	237		1631	39	26873	379039	70.90
70	900	12392	7454	3454	683	212		1499	36	26630	352166	75.62
71	1066	12427	7116	3369	650	187		1376	33	26234	325536	80.59
72	1258	12348	6722	3273	615	165		1252	32	25675	299302	85.78
73	1462	12152	6291	3166	580	142		1136	29	24958	273627	91.21
74	1677	11857	5828	3048	544	123		1019	27	24113	248669	96.97
75	1897	11467	5348	2919	509	105		902	25	23172	224556	103.19
76	2087	10911	4855	2785	473	89		801	23	22024	201384	109.36
77	2261	10317	4373	2639	439	75		706	20	20829	179360	116.13
78	2397	9855	3885	2488	404	63		612	17	19531	158531	123.20
79	2485	8942	3419	2336	371	51		532	14	18150	139000	130.58
80	2518	8197	2976	2185	340	41		458	13	16728	120850	138.42
81	2493	7430	2567	2023	308	33		400	12	15266	104122	146.62
82	2453	6658	2185	1869	255	27		327	10	13784	88856	155.13
83	2285	5900	1841	1715	251	19		277	9	12297	75072	163.80
84	2111	5182	1531	1567	223	16		233	7	10870	62775	173.16
85	1904	4479	1262	1416	206	12		189	7	9475	51905	182.55
86	1681	3828	1022	1275	176	9		167	6	8164	42430	192.41
87	1444	3233	818	1139	154	5		131	4	6928	34266	202.18
88	1213	2689	647	1010	133	3		102	4	5801	27338	212.20
89	992	2208	502	893	114	2		80	3	4794	21537	222.59
90	795	1786	389	782	97			65	3	3917	16743	233.95
91	617	1414	292	678	82			51	3	3137	12826	244.58
92	468	1103	217	579	68			43	1	2479	9689	255.86
93	345	838	163	496	56			29		1927	7210	267.27
94	248	623	118	418	45			22		1474	5283	279.01
95	174	446	87	349	34			22		1112	3809	291.94
96	124	308	63	286	27			14		822	2697	304.78
97	79	196	45	231	21			14		586	1875	312.53
98	53	118	35	187	16			9		418	1289	324 20

Age	I	II	III	IV	V	VI	VIIa	VIIb	VIIIa	VIIIb	dx	lx	1000qx
99	44	58	28	143	9			6			288	871	330.65
100	22	32	22	113	5			3			197	583	337.91
101	15	21	15	81	1			1			134	386	347.14
102	12	15	11	63							101	252	400.79
103	9	11	9	35							64	151	423.85
104	6	7	7	20							40	87	459.76
105	5	6	6	6							23	47	489.31
106	4	4	5								13	24	541.66
107	2	3	2								7	11	636.36
108	1	1	1								3	4	731.71
109	1										1	1	1000.00

Mortality Table

Massachusetts Males 1914—1916. (Series C)

Age	I	II	III	IV	V	VI	VIIa	VIIb	VIIIa	VIIIb	dx	lx	1000qx
10			24	140	268	126	34		1780	5	2377	1000000	2.38
11			33	156	301	246	98		1615	6	2455	997623	2.46
12			41	175	337	295	189		1455	8	2500	995168	2.51
13			53	195	374	352	357		1308	10	2649	992668	2.67
14			64	218	414	414	453	2	1197	11	2773	990019	2.80
15			78	242	456	484	562	2	1086	14	2924	987246	2.96
16			94	268	500	560	628	2	991	16	3059	984322	3.10
17			113	295	544	639	693	2	899	19	3204	981263	3.27
18			136	324	597	728	744	3	809	22	3363	978059	3.44
19			161	355	648	823	790	3	729	26	3535	974696	3.63
20		21	188	388	700	920	825	5	647	29	3723	971161	3.83
21		53	218	423	752	1023	850	7	570	32	3928	967438	4.06
22		64	254	460	807	1127	869	9	495	37	4122	963510	4.28
23		80	291	499	862	1234	883	10	418	40	4317	959388	4.50
24		101	330	539	918	1348	880	14	345	45	4520	955071	4.73
25		127	371	581	974	1462	856	17	281	50	4719	950551	4.96
26		154	415	627	1031	1574	820	21	220	55	4917	945832	5.20
27		184	463	672	1086	1684	792	27	170	59	5137	940915	5.46
28		219	507	720	1145	1793	756	33	133	64	5370	935776	5.74
29		262	552	772	1200	1900	718	39	103	69	5615	930408	6.03
30		311	596	822	1256	2002	688	48	77	75	5875	924793	6.35
31		361	641	879	1309	2099	657	58	58	80	6142	918918	6.68
32		422	682	933	1362	2191	614	69	45	85	6413	912776	7.03

33	486	722	995	1413	2278	592	82	32	90	6590	906383	7.38
34	553	761	1056	1464	2357	562	96	22	95	6966	899673	7.74
35	629	797	1119	1511	2428	533	111	10	100	7238	892707	8.11
36	708	834	1185	1558	2490	504	130	6	104	7514	885469	8.49
37	783	868	1252	1605	2546	480	149	3	109	7795	877955	8.88
38	870	906	1326	1647	2594	448	171		112	8074	870160	9.28
39	954	947	1400	1683	2633	422	194		117	8350	862086	9.69
40	1046	997	1477	1728	2663	398	219		120	8648	853736	10.13
41	1136	1057	1558	1764	2685	374	245		124	8943	845088	10.58
42	1230	1125	1640	1798	2698	351	271		125	9238	836145	11.05
43	1329	1211	1726	1829	2704	324	300		128	9551	826907	11.55
44	1426	1319	1828	1860	2702	300	329		130	9894	817356	12.10
45	1534	1452	1908	1887	2680	278	358		133	10239	807462	12.68
46	1657	1606	2002	1910	2671	255	387		135	10635	797223	13.34
47	1780	1799	2098	1932	2444	233	417		135	11057	786588	14.05
48	1924	2025	2204	1949	2609	213	446		137	11383	775531	14.87
49	2089	2287	2294	1965	2568	192	472		137	12041	764098	15.75
50	2273	2577	2393	1977	2519	173	497		137	12597	751957	16.75
51	2490	2900	2495	1983	2462	154	521		137	13213	739360	17.87
52	2741	3258	2595	1990	2400	139	542		135	13892	726147	19.13
53	3022	3644	2693	1991	2330	122	559		135	14616	712255	20.52
54	3345	4049	2790	1988	2254	105	574		133	15391	697639	22.06
55	3711	4469	2885	1980	2173	93	586		132	16221	682248	23.78
56	4116	4890	2972	1969	2084	81	593		130	17069	666027	25.62
57	4559	5311	3059	1955	1993	72	597		127	17955	648958	27.66
58	5033	5722	3137	1933	1898	63	597		125	18845	631003	29.86
59	5545	6116	3207	1910	1798	53	592		122	19732	612158	32.23
60	6080	6467	3272	1880	1695	45	585		119	20588	592426	34.75
61	6660	6786	3327	1846	1589	36	573		116	21438	571838	37.49
62	7188	7050	3373	1807	1484	22	557		111	22157	550400	40.26
63	7715	7257	3406	1765	1379	12	538		108	22818	528243	43.20
64	8235	7400	3432	1718	1275	5	518		103	23405	505425	46.31
65	8720	7477	3443	1669	1170	2	494		98	23888	482020	49.56
66	9157	7484	3442	1614	1069		468		93	24263	458132	52.96
67	9533	7416	3432	1553	969		430		88	24530	433869	56.54
68	9835	7284	3405	1494	876		413		84	24667	409339	60.26
69	10054	7087	3367	1430	782		384		79	24700	384672	64.21
70	10180	6828	3317	1365	697		353		74	24619	359972	68.39
71	10208	6520	3254	1298	615		324		69	24428	335353	72.84
72	10143	6159	3130	1228	541		295		64	24131	310925	77.61
73	9982	5763	3094	1156	469		267		59	23723	286794	82.72
74	9740	5339	2998	1086	405		240		55	23226	263071	88.29

Age	I	II	III	IV	V	VI	VIIa	VIIb	VIIIa	VIIIb	dx	lx	1000qx
75	3804	9430	4900	2893	1016	346		213		50	22652	239845	94.44
76	4176	8963	4448	2778	946	295		189		45	21850	217193	100.61
77	4534	8475	4005	2658	876	247		166		40	21001	195343	107.54
78	4807	7931	3560	2530	807	206		144		37	20022	174342	114.84
79	4983	7346	3132	2399	742	168		125		32	18927	154320	122.65
80	5050	6738	2726	2262	678	136		108		29	17722	135393	130.89
81	4999	6103	2351	2122	615	109		94		26	16419	117671	139.53
82	4919	5469	2001	1982	508	87		77		22	15065	101252	148.79
83	4583	4847	1686	1842	500	65		65		19	13607	86187	157.88
84	4234	4257	1403	1703	446	51		55		16	12165	72580	167.61
85	3820	3679	1156	1562	411	37		45		14	10724	60415	177.51
86	3371	3145	937	1429	349	26		39		11	9307	49691	187.30
87	2897	2656	749	1298	305	17		31		10	7963	40384	197.18
88	2432	2209	593	1172	265	10		24		8	6713	32421	207.06
89	1990	1813	460	1053	227	6		19		6	5574	25708	216.82
90	1594	1467	355	937	195	1		15		5	4569	20134	226.93
91	1237	1161	267	831	164			12		5	3677	15565	236.23
92	938	905	199	729	136			10		2	2919	11888	245.54
93	692	689	149	636	112			7			2285	8969	254.77
94	497	511	109	549	89			5			1760	6684	263.32
95	349	340	80	468	70			3			1312	4924	266.45
96	252	234	57	397	55			3			998	3612	276.30
97	158	149	41	331	41			3			721	2614	275.82
98	106	90	32	272	30			2			532	1893	281.04
99	89	44	25	220	19			2			399	1361	293.17
100	44	25	19	174	11			2			275	962	285.86
101	30	16	17	134	3						202	687	294.03
102	24	12	17	100							153	485	315.46
103	18	9	15	70							112	332	337.35
104	12	5	15	45							77	220	350.00
105	10	5	15	24							54	143	377.62
106	8	3	14	7							32	89	359.55
107	4	2	13								19	57	333.33
108	2	1	11								14	38	369.42
109	2		11								13	24	541.67
110			11								11	11	1000.00

Mortality Table

Michigan Males 1909—1915

Age	I	II	III	IV	V	VI	VIIa	VIIb	VIII	dx	lx	1000qx
10			25	135	149	176	93		1822	2402	1000000	2.40
11			34	150	168	207	172		1888	2621	997598	2.63
12			44	168	187	241	260		1920	2822	994977	2.84
13			54	187	207	279	353		1947	3029	992155	3.06
14			65	207	229	319	457	1	1877	3157	989126	3.19
15			79	227	252	362	570	1	1736	3229	985969	3.27
16			95	249	278	409	688	1	1532	3254	982740	3.31
17			110	273	303	456	809	1	1344	3298	979486	3.37
18			134	296	330	509	929	2	1136	3338	976188	3.42
19		29	155	321	357	564	1041	2	938	3409	972850	3.50
20		37	179	344	387	618	1146	4	759	3476	969441	3.59
21		46	208	370	417	675	1238	4	603	3563	965965	3.69
22		60	238	396	447	736	1316	5	471	3669	962402	3.81
23		74	270	422	479	797	1378	7	362	3789	958733	3.95
24		89	305	448	511	856	1421	9	274	3913	954944	4.10
25		111	339	474	544	917	1448	12	205	4050	951031	4.26
26		134	376	500	577	968	1456	13	151	4175	946981	4.41
27		163	414	526	611	1036	1446	18	110	4324	942806	4.59
28		194	450	551	644	1093	1423	22	80	4457	938482	4.75
29		234	487	576	677	1149	1382	26	57	4588	934025	4.91
30		280	520	603	711	1198	1331	31	40	4714	929437	5.07
31		331	553	628	744	1246	1270	39	27	4838	924723	5.23
32		388	582	654	777	1299	1201	45	19	4965	919885	5.40
33		451	608	679	810	1345	1125	55	14	5086	914920	5.56
34		523	630	707	842	1383	1044	64	9	5202	909834	5.72
35		597	651	734	873	1422	963	74	6	5320	904632	5.88
36		683	666	759	905	1456	880	86	3	5438	899312	6.05
37		771	677	795	935	1485	799	99	2	5563	893874	6.22
38		866	691	828	964	1509	719	113	2	5692	888311	6.41
39		977	704	864	990	1529	644	128	1	5837	882619	6.61
40	14	1086	716	906	1023	1547	573	145	1	6011	876782	6.86
41	25	1205	736	948	1050	1560	506	162	1	6193	870771	7.11
42	35	1331	758	997	1081	1568	443	179		6392	864578	7.39
43	46	1465	795	1049	1103	1571	387	199		6615	858186	7.71
44	61	1605	843	1106	1129	1572	335	217		6868	851571	8.07
45	80	1757	918	1169	1152	1567	290	237		7168	844703	8.49
46	102	1920	1000	1235	1178	1559	248	256		7498	837535	8.95
47	130	2094	1114	1309	1201	1549	211	276		7884	830037	9.50
48	161	2282	1253	1386	1224	1533	180	295		8314	822153	10.11

Age	I	II	III	IV	V	VI	VIIa	VIIb	VIII	dx	lx	1000qx
49	194	2491	1427	1470	1245	1514	151	312		8804	813839	10.82
50	235	2717	1632	1557	1266	1490	127	329		9353	805035	11.62
51	277	2966	1834	1649	1285	1464	106	345		9926	795682	12.47
52	324	3240	2138	1744	1303	1436	89	359		10633	785756	13.53
53	371	3543	2440	1841	1322	1403	72	369		11361	775123	14.66
54	420	3777	2768	1939	1338	1367	60	380		12049	763762	15.78
55	467	4237	3131	2037	1352	1326	49	388		12987	751713	17.28
56	512	4640	3500	2134	1364	1287	40	392		13869	738726	18.77
57	554	5074	3887	2229	1375	1241	32	394		14786	724857	20.40
58	592	5543	4280	2320	1384	1188	27	394		15728	710071	22.15
59	627	6048	4677	2407	1391	1143	22	391		16706	694343	24.06
60	663	6580	5063	2485	1395	1095	17	386		17684	677637	26.10
61	701	7134	5426	2557	1397	1043	14	379		18651	659953	28.26
62	744	7702	5763	2619	1396	990	11	368		19593	641302	30.55
63	804	8282	6067	2671	1393	934	8	356		20515	621709	33.00
64	881	8860	6327	2712	1387	880	8	342		21397	601194	35.59
65	988	9426	6531	2740	1378	856	5	326		22250	579797	38.38
66	1133	9974	6683	2754	1363	769	4	309		22989	557547	41.23
67	1321	10489	6774	2755	1349	713	4	290		23695	534558	44.33
68	1563	10920	6803	2744	1331	658	3	273		24295	510863	47.56
69	1858	11354	6768	2717	1312	605	2	254		24870	486568	53.17
70	2212	11691	6672	2676	1283	555	1	234		25324	461698	54.85
71	2617	11948	6517	2622	1256	504	1	214		25679	436374	58.85
72	3064	12120	6303	2557	1225	456	1	195		25921	410695	63.11
73	3543	12228	6040	2478	1190	411	1	177		26068	384774	67.75
74	4039	12185	5738	2390	1154	367		159		26032	358706	72.57
75	4539	12071	5398	2288	1114	327		141		25878	332674	77.79
76	5009	11860	5035	2179	1073	290		125		25571	306796	83.35
77	5415	11557	4639	2067	1030	253		110		25071	281225	89.15
78	5799	11174	4241	1946	985	221		95		24461	256154	95.49
79	6078	10709	3838	1818	938	191		83		23655	231693	102.10
80	6258	10174	3381	1689	889	164		71		22626	208038	108.35
81	6321	9586	3048	1559	841	140		62		21557	185412	116.27
82	6278	8952	2670	1426	793	119		51		20289	163855	123.82
83	6118	8275	2316	1302	743	99		43		18896	143566	131.62
84	5862	7588	1992	1171	694	83		36		17426	124670	139.78
85	5511	6897	1691	1049	645	67		30		15890	107244	148.17
86	5092	6208	1419	932	597	55		26		14329	91354	156.85
87	4619	5529	1181	817	549	44		21		12760	77025	165.66
88	4111	4880	970	708	504	34		16		11223	64265	174.64
89	3596	4263	789	608	459	26		13		9754	53042	183.89
90	3077	3686	633	515	417	20		10		8358	43288	193.08
91	2592	3140	503	430	376	15		8		7064	34930	202.23

92	2136	2651	396	353	338	10	7	5891	27866	211.40
93	1729	2208	310	281	301	7	4	4840	21975	220.25
94	1372	1814	222	219	267	4	4	3902	17135	227.72
95	1070	1474	186	162	233	2	4	3131	13233	236.61
96	815	1174	146	114	204		2	2455	10102	248.02
97	610	920	114	73	176		2	1895	7647	247.81
98	451	706	91	36	150		1	1435	5752	249.48
99	331	526	78	6	126		1	1068	4317	247.39
100	241	383	65		105		1	795	3249	244.69
101	175	263	60		87		1	586	2454	238.79
102	130	171	54		70			425	1868	227.52
103	100	100	50		54			304	1443	210.67
104	78	51	48		40			217	1139	190.52
105	66	14	46		28			154	922	167.03
106	55		45		19			119	768	154.95
107	48		41		11			100	649	154.08
108	42		40		3			85	549	154.83
109	38		38					76	464	163.79
110	35		36					71	388	182.99
111	31		32					63	317	198.74
112	26		30					56	254	220.47
113	21		28					49	198	247.47
114	15		24					39	149	261.74
115	10		24					34	110	309.09
116	4		18					22	76	289.47
117	1		16					17	54	314.81
118			16					16	37	432.43
119			12					12	21	571.43
120			9					9	9	1000.00

Mortality Table

Massachusetts Males 1914—1916. (Series B)

Age	I	II	III	IV	V	VI	VIIa	VIIb	dx	lx	1000qx
10			42	81	228	258	2627	36	3272	1000000	3.27
11			50	93	252	309	2436	44	3186	996728	3.20
12			60	106	289	367	2246	53	3121	993542	3.14
13		28	74	122	322	429	2052	63	3090	990421	3.12
14		39	87	139	357	498	1862	74	3056	987331	3.10
15		48	106	158	395	572	1677	86	3042	984275	3.09
16		56	122	180	435	651	1501	99	3044	981233	3.10
17		69	143	201	478	739	1335	114	3079	978189	3.15

Age	I	II	III	IV	V	VI	VIIa	VIIb	dx	lx	1000qx
18		84	168	227	520	830	1181	129	3139	975110	3.22
19		102	193	252	565	924	1039	146	3221	971971	3.31
20		121	223	282	612	1024	910	163	3335	968750	3.44
21		145	253	313	660	1125	792	181	3470	965415	3.59
22		173	288	347	708	1232	687	200	3635	961945	3.78
23		206	324	382	758	1339	593	219	3821	958310	3.99
24		238	363	419	808	1447	509	239	4023	954489	4.21
25		279	404	461	857	1554	436	260	4251	950466	4.47
26		320	448	502	908	1660	372	280	4490	946215	4.75
27		368	490	547	959	1766	316	301	4747	941725	5.04
28		422	535	595	1008	1866	267	321	5014	936978	5.35
29		478	581	646	1059	1963	225	342	5294	931964	5.68
30		537	627	699	1106	2055	189	362	5575	926670	6.02
31		604	674	754	1153	2145	158	381	5867	921095	6.37
32		671	719	812	1199	2222	132	399	6154	915228	6.72
33		742	763	874	1244	2295	110	416	6444	909074	7.09
34		818	811	938	1288	2358	91	433	6737	902630	7.41
35		894	861	1006	1329	2415	73	448	7026	895803	7.84
36	6	974	905	1075	1370	2464	62	462	7316	888867	8.23
37	10	1054	956	1150	1407	2503	52	475	7607	881549	8.63
38	13	1141	1012	1226	1444	2534	42	486	7898	873942	9.04
39	18	1232	1068	1306	1478	2556	35	496	8189	866044	9.46
40	25	1323	1134	1391	1520	2569	28	504	8493	857855	9.90
41	32	1416	1210	1478	1541	2573	23	511	8784	849362	10.34
42	44	1519	1293	1572	1571	2571	18	517	9105	840578	10.83
43	56	1626	1389	1666	1594	2559	13	521	9424	831473	11.33
44	73	1742	1498	1765	1619	2540	12	523	9772	822049	11.89
45	93	1872	1617	1867	1641	2514	10	524	10138	812277	12.48
46	120	2011	1763	1972	1669	2481	8	524	10538	802139	13.14
47	151	2165	1926	2079	1677	2442	7	522	10969	791601	13.86
48	185	2342	2115	2189	1698	2398	5	519	11451	780632	14.67
49	227	2532	2324	2301	1703	2347	4	515	11954	769181	15.54
50	279	2747	2554	2415	1713	2291	3	510	12512	757227	16.52
51	333	2991	2807	2528	1720	2230	3	504	13116	744715	17.61
52	399	3262	3083	2639	1731	2165	2	497	13778	731599	18.83
53	471	3554	3376	2752	1725	2097	2	489	14466	717821	20.15
54	553	3879	3685	2860	1723	2027	1	479	15207	703355	21.62
55	643	4227	4009	2964	1717	1951	1	469	15981	688148	23.22
56	742	4608	4333	3064	1708	1873	1	458	16787	672167	24.97
57	854	5009	4664	3158	1696	1792		446	17619	655380	26.88
58	974	5426	4992	3244	1677	1711		434	18458	637761	28.94
59	1105	5857	5315	3321	1660	1627		420	19305	619303	31.17
60	1247	6301	5618	3389	1638	1543		406	20142	599998	33.57

61	1405	6740	5899	3444	1611	1457
62	1573	7178	6148	3488	1580	1336
63	1758	7597	6369	3520	1546	1218
64	1946	7994	6549	3537	1508	1202
65	2153	8364	6686	3541	1468	1118
66	2371	8686	6775	3528	1422	1035
67	2601	8961	6816	3503	1375	955
68	2833	9186	6807	3462	1325	878
69	3070	9349	6745	3406	1273	803
70	3304	9443	6633	3337	1218	731
71	3528	9466	6474	3253	1162	660
72	3736	9425	6270	3155	1104	596
73	3931	9312	6027	3047	1044	534
74	4093	9126	5745	2925	984	477
75	4215	8881	5436	2799	923	423
76	4296	8567	5101	2661	864	372
77	4336	8206	4750	2519	804	326
78	4325	7797	4388	2371	743	284
79	4261	7349	4023	2221	686	245
80	4164	6859	3638	2066	630	212
81	3984	6359	3284	1912	575	181
82	3780	5849	2928	1761	522	154
83	3538	5327	2585	1608	472	130
84	3262	4801	2261	1464	422	109
85	2970	4301	1962	1321	378	90
86	2660	3805	1685	1186	335	74
87	2347	3342	1430	1054	296	60
88	2038	2896	1201	931	259	48
89	1743	2491	995	817	220	38
90	1461	2331	816	711	193	30
91	1201	1777	658	614	165	23
92	970	1465	522	524	138	18
93	764	1199	406	442	116	14
94	589	961	308	367	94	10
95	437	755	226	303	76	8
96	315	582	164	246	61	5
97	219	435	113	194	46	3
98	141	314	72	152	34	2
99	82	216	41	114	24	1
100	41	136	17	82	16	1
101	11	78	2	55	9	
102		32		34	2	
103				16		
104				2		

391	20947	579856	36.12
376	21679	558909	38.79
360	22362	537230	41.62
344	23080	514868	44.83
327	23657	491788	48.10
310	24127	468131	51.54
293	24504	444004	55.19
276	24767	419500	59.04
259	24906	394733	63.09
242	24908	369828	67.35
211	24754	344920	71.77
208	24494	320166	76.50
192	24087	295672	81.47
176	23526	271585	86.62
161	22838	248059	92.07
146	22007	225221	97.71
131	21072	203214	103.69
118	20026	182142	109.95
106	18891	162116	116.53
94	17645	143225	123.20
83	16373	125580	130.42
73	15067	109202	137.97
64	13724	94135	145.79
55	12374	80411	153.88
48	11070	68037	162.71
41	9786	56967	171.78
34	8563	47181	181.49
29	7402	38618	191.67
24	6328	31216	202.71
20	5562	24888	223.48
16	4454	19325	230.47
13	3650	14872	245.43
10	2951	11222	262.97
8	2336	8271	282.43
6	1811	5935	305.14
5	1378	4124	334.14
3	1013	2746	368.90
2	717	1733	413.73
2	480	1016	472.44
1	294	536	548.51
1	156	242	644.63
	68	86	790.70
	16	18	888.89
	2	2	1000.00

ADDENDA II

In order to show a rapid application of frequency curve methods to the graduation of mortality tables when the number of lives exposed to risk at various ages is known, the following data, relating to applicants who had been rejected for life assurance on account of impaired health, by Scandinavian assurance companies is instructive. The original statistics as collected by a committee of the insurance companies were first published in the quinquennial report (1910—1915) of the Danish Government Life Assurance Institution (The Statsanstalt) for 1917.

The material related to Scandinavian and Finnish applicants who previously to 1893 (and in the case of two Danish companies before 1899) had been rejected for life assurance. By a special investigation, the committee followed up these rejections and sought to establish whether the applicants were alive at July 1, 1899, or were previously deceased. Detailed reports for the full period during which the risks were under observation were available for 8,208 individual applicants. For 2,023 applicants complete data were not available.

The final statistical results of the Statsanstalt's investigation are shown in the following summary table:

TABLE I.

Mortuary Experience of Rejected Risks of Scandinavian Life Companies.

Attained Age	No. Exposed to Risk	Number of Deaths
15-19	434	6
20-24	3,831	28
25-29	11,405	145
30-34	17,644	233
35-39	19,442	318
40-44	17,600	324
45-49	13,971	296
50-54	10,179	295
55-59	6,640	264
60-64	3,927	194
65-69	1,995	96
70-74	836	71
75-79	306	32
80-84	98	20
85-89	12	3

The exposed to risk by separate ages and the correlated deaths are shown in Table II in Columns 2 and 3, from which we, without difficulty, obtain the crude or ungraduated mortality rates, as shown Column 4.

We next assume a purely hypothetical frequency distribution of the exposed to risk, according to age, represented by a Laplacean normal probability curve with its mean or origin at age fifty and a dispersion equal to 12.5 years, as shown in Column 5. The frequency distribution of the number of deaths on the basis of the ungraduated mortality rates in Column 4

and the above-mentioned normal probability curve is shown in Column 6, which may be considered as an ungraduated compound frequency curve. [1]

Arranged in quinquennial age intervals this latter frequency distribution is shown in the following summary table:

Ages	No. of Deaths
13-17	51
18-22	75
23-27	329
28-32	711
33-37	1,464
38-42	2,498
43-47	3,649
48-52	5,377
53-57	6,238
58-62	6,232
63-67	5,254
68-72	3,605
73-77	2,536
78-82	1,425
83-87	1,169
88-92	351
93 or over	95
Total...	41,059

The above frequency distribution is now subjected to a graduation by means of the Laplacean—Charlier or Gram—Charlier frequency function. The mathematical calculations give the following parameters:

[1] A slight adjustment was made in the figures in column (6) corresponding to age 70, and in the age groups above the age of 88.

Mean Age	57.75 years
Dispersion	13.32 years
Skewness	—0.0031
Excess	—0.0037

Applying these parameters to standard probability tables we obtain the usual Laplacean—Charlier frequency curve. Distributing the 41,059 individual deaths according to this frequency curve we obtain column (7) which is the graduated death curve corresponding to the hypothetical exposure as given by column (5). The final mortality rates per 1,000 of exposed to risk are then found by dividing (7) with (5) and are shown in column (8).

In order to show how close the graduation by means of frequency curves agrees with the actual observations, I have made a calculation of the "actual" to the "expected" deaths by quinquennial age intervals as shown in the following table:

TABLE III.

Comparison between "Actual" and "Expected" Deaths on the Basis of the Graduated Mortality Rates of the Scandinavian Mortality Table for Rejected Lives

Ages	No. Exposed to Risk	Actual Deaths	Expected Deaths
15-19	434	6	3.4
20-24	3,831	28	37.6
25-29	11,405	145	133.4
30-34	17,644	233	242.2
35-39	19,442	318	314.3
40-44	17,600	324	336.8

Ages	No. Exposed to Risk	Actual Deaths	Expected Deaths
45-49	13,971	296	321.8
50-54	10,179	295	287.2
55-59	6,640	264	234.8
60-64	3,927	194	178.6
65-69	1,995	96	119.5
70-74	836	71	67.4
75-79	306	32	33.8
80-84	98	20	15.1
85-89	12	3	2.5
Total	108,320	2,325	2,328.4

Considering the somewhat meager experience on which the graduation was based, I think it must be admitted that the method of frequency curves comes surprisingly close to the actual facts. In this connection it is of interest to note that the actuaries of the Danish Statsanstalt made a graduation of the above data on the basis of Makeham's method and obtained from least square methods the following values for the constants. [1]

$$A = 0.006$$
$$\log B = 7.0566 - 10$$
$$\log C = 0.025$$

The "expected" deaths according to this latter graduation, and on the basis of the above experience, amount in total to 2,317 as against 2,325 "actual" deaths and 2,328 "expected" deaths according to the frequency curve method. Viewed from the stand-

[1] See formula (6) page 192 of Institute of Actuaries Text Book. **Life Contingencies** by E. F. Spurgeon, London, 1922.

point of the principle of least squares it is also found that the sum of the squares of the deviations is smaller under the frequency curve method than under the method of Makeham, which seems to be pretty good evidence of the soundness of the method in spite of the fact that I throughout have worked with unweighted observations. If properly chosen weights were applied to the observations even closer results could be obtained.

TABLE II.

Mortality Experience of Rejected Scandinavian Risks (Male).

(1) Age	(2) Exposed to Risk	(3) No. of Deaths	(4) (3) : (2)	(5) Hypothetical Exposure	(6) (5) × (4) Crude Death Curve	(7) Graduated Death Curve	(8) (7) : (5) $1000q_x$
15	11	0	0.00000	792	0	5.6	7.07
16	31	1	0.03226	987	32	7.1	7.07
17	64	1	0.01562	1223	19	9.2	7.52
18	121	0	0.00000	1506	0	11.7	7.77
19	207	4	0.01932	1842	3	15.4	8.36
20	340	1	0.00294	2239	7	19.7	8.80
21	501	1	0.00200	2705	5	25.0	9.24
22	719	6	0.00834	3246	27	30.8	9.49
23	982	6	0.00611	3871	24	38.8	10.02
24	1289	14	0.01086	4586	50	47.8	10.42
25	1619	22	0.01359	5399	73	58.2	10.78
26	1986	23	0.01158	6316	73	70.6	11.18
27	2287	34	0.01487	7341	109	85.0	11.58
28	2597	29	0.01117	8478	95	101.7	12.00
29	2916	37	0.01269	9728	123	120.5	12.39
30	3180	38	0.01195	11092	133	142.0	12.80
31	3395	50	0.01473	12566	185	166.4	13.24
32	3564	44	0.01235	14146	175	193.5	13.68
33	3700	46	0.01243	15822	197	223.4	14.12
34	3806	55	0.01445	17585	254	257.0	14.61
35	3882	48	0.01236	19419	240	293.3	15.10
36	3943	64	0.01623	21307	346	332.8	15.62
37	3921	72	0.01836	23230	427	375.3	16.16
38	3880	66	0.01701	25164	428	420.0	16.69
39	3816	68	0.01782	27086	483	467.7	17.27
40	3737	66	0.01766	28969	512	517.6	17.87
41	3637	63	0.01732	30785	533	566.9	18.41

(1) Age	(2) Exposed to Risk	(3) No. of Deaths	(4) (3) : (2)	(5) Hypothetical Exposure	(6) (5) × (4) Crude Death Curve	(7) Graduated Death Curve	(8) (7) : (5) 1000qx
42	3539	59	0.01667	32506	542	623.3	19.17
43	3426	62	0.01810	34105	617	678.2	19.89
44	3261	74	0.02269	35553	807	732.7	20.61
45	3079	67	0.02176	36827	801	787.8	21.39
46	2941	61	0.02074	37903	786	842.4	22.23
47	2793	46	0.01647	38762	638	895.1	22.97
48	2653	61	0.02299	39387	906	945.9	24.02
49	2505	61	0.02435	39767	968	994.3	25.00
50	2348	61	0.02598	39894	1036	1039.0	26.04
51	2184	65	0.02976	39767	1183	1079.9	27.16
52	2024	66	0.03261	39387	1284	1116.0	28.33
53	1882	59	0.03135	38762	1215	1147.4	29.53
54	1741	44	0.02527	37903	958	1173.3	30.96
55	1610	62	0.03851	36827	1418	1193.0	32.39
56	1447	60	0.04147	35553	1474	1206.9	33.95
57	1308	45	0.03440	34105	1173	1214.3	35.60
58	1189	47	0.03953	32506	1285	1214.9	37.37
59	1086	50	0.04604	30785	1417	1209.0	39.27
60	966	44	0.04555	28969	1320	1197.0	41.32
61	871	35	0.04019	27186	1089	1178.8	43.52
62	786	35	0.04453	25164	1121	1154.2	45.87
63	701	44	0.06277	23230	1458	1124.6	48.41
64	603	36	0.05970	21307	1272	1090.1	51.16
65	518	22	0.04247	19419	825	1050.7	54.11
66	453	24	0.05298	17585	932	1006.3	57.22
67	392	19	0.04847	15822	767	960.1	60.68
68	340	16	0.04706	14146	666	909.6	64.30
69	291	15	0.05155	12566	648	858.4	68.31
70	244	25	0.10246	11092	1136	804.2	72.50
71	193	17	0.08808	9728	857	750.9	77.19
72	158	13	0.08228	8478	698	695.7	82.06
73	132	9	0.06818	7341	501	642.4	87.51
74	109	7	0.06422	6316	406	589.1	93.27
75	91	8	0.08791	5399	475	537.7	99.59
76	74	10	0.13514	4586	620	486.8	106.15
77	58	8	0.13793	3871	534	440.3	113.74
78	45	4	0.08889	3246	289	393.8	121.32
79	37	2	0.05405	2705	146	351.9	130.09
80	31	5	0.16129	2239	361	311.8	139.26
81	24	6	0.25000	1842	461	274.5	149.02
82	18	2	0.11112	1506	168	241.6	160.42
83	15	4	0.26667	1223	326	209.5	171.30
84	9	3	0.33334	987	329	181.5	183.89
85	6	2	0.33334	792	264	155.9	196.84
86	3	0	0.00000	631	000	133.4	211.41
87	2	1	0.50000	499	250	113.4	227.26
88	2	1	0.50000	393	197	95.5	243.00
89	0.5	0	0.50000	307	154	79.2	257.98

Note:—The observations above age 87 are not reliable.

TABLE OF CONTENTS

CHAPTER I.

Introduction to the Theory of Frequency Curves.

CHAPTER II.

The Human Death Curve.

INDEX